PATENT COOPERATION TREATY

An Dr. Krauß zur Kenntnis! 27.2.) 26.3.01

From the INTERNATIONAL SEARCHING AUTHORITY

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT OR THE DECLARATION

(PCT Rule 44.1)

To:
BOEHMERT & BOEHMERT
Attn. GODDAR, Heinz
Hollerallee 32
D-28209 Bremen
GERMANY

Boehmert & Boehmert
Bremen
Eing. 26. MRZ. 2001
Frist 26.5.01

Applicant's or agent's file reference L10046 PCT	Date of mailing *(day/month/year)* 26/03/2001 **FOR FURTHER ACTION** See paragraphs 1 and 4 below
International application No. PCT/EP 00/ 09363	International filing date *(day/month/year)* 26/09/2000
Applicant LION BIOSCIENCE AG	

1. [X] The applicant is hereby notified that the International Search Report has been established and is transmitted herewith.

Filing of amendments and statement under Article 19:
The applicant is entitled, if he so wishes, to amend the claims of the International Application (see Rule 46):

When? The time limit for filing such amendments is normally 2 months from the date of transmittal of the International Search Report; however, for more details, see the notes on the accompanying sheet.

Where? Directly to the International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland
Fascimile No.: (41-22) 740.14.35

For more detailed instructions, see the notes on the accompanying sheet.

2. [ ] The applicant is hereby notified that no International Search Report will be established and that the declaration under Article 17(2)(a) to that effect is transmitted herewith.

3. [ ] **With regard to the protest** against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:

[ ] the protest together with the decision thereon has been transmitted to the International Bureau together with the applicant's request to forward the texts of both the protest and the decision thereon to the designated Offices.

[ ] no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. **Further action(s):** The applicant is reminded of the following:

Shortly after **18 months** from the priority date, the international application will be published by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau as provided in Rules 90*bis*.1 and 90*bis*.3, respectively, before the completion of the technical preparations for international publication.

Within **19 months** from the priority date, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase until 30 months from the priority date (in some Offices even later).

Within **20 months** from the priority date, the applicant must perform the prescribed acts for entry into the national phase before all designated Offices which have not been elected in the demand or in a later election within 19 months from the priority date or could not be elected because they are not bound by Chapter II.

Name and mailing address of the International Searching Authority European Patent Office, P.B. 5818 Patentlaan 2 NL-2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Lucia Van Pinxteren

Form PCT/ISA/220 (July 1998)

PATENT COOPERATION TREATY

From the
INTERNATIONAL PRELIMINARY EXAMINING AUTHORITY

To:

GODDAR, Heinz
BOEHMERT & BOEHMERT
Hollerallee 32
28209 Bremen
ALLEMAGNE

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL PRELIMINARY EXAMINATION REPORT
(PCT Rule 71.1)

Applicant's or agent's file reference L10046 PCT Alte bei Dr. Schabe (Ktw)	Date of mailing *(day/month/year)* 27.12.2001 **IMPORTANT NOTIFICATION**

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/EP00/09363	26/09/2000	01/10/1999

Applicant
LION BIOSCIENCE AG

1. The applicant is hereby notified that this International Preliminary Examining Authority transmits herewith the international preliminary examination report and its annexes, if any, established on the international application.

2. A copy of the report and its annexes, if any, is being transmitted to the International Bureau for communication to all the elected Offices.

3. Where required by any of the elected Offices, the International Bureau will prepare an English translation of the report (but not of any annexes) and will transmit such translation to those Offices.

4. **REMINDER**

The applicant must enter the national phase before each elected Office by performing certain acts (filing translations and paying national fees) within 30 months from the priority date (or later in some Offices) (Article 39(1)) (see also the reminder sent by the International Bureau with Form PCT/IB/301).

Where a translation of the international application must be furnished to an elected Office, that translation must contain a translation of any annexes to the international preliminary examination report. It is the applicant's responsibility to prepare and furnish such translation directly to each elected Office concerned.

For further details on the applicable time limits and requirements of the elected Offices, see Volume II of the PCT Applicant's Guide.

Name and mailing address of the IPEA/	Authorized officer
European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	Hingel, W Tel.+49 89 2399-8717

Form PCT/IPEA/416 (July 1992)

NOTES TO FORM PCT/ISA/220

These Notes are intended to give the basic instructions concerning the filing of amendments under article 19. The Notes are based on the requirements of the Patent Cooperation Treaty, the Regulations and the Administrative Instructions under that Treaty. In case of discrepancy between these Notes and those requirements, the latter are applicable. For more detailed information, see also the PCT Applicant's Guide, a publication of WIPO.

In these Notes, "Article", "Rule", and "Section" refer to the provisions of the PCT, the PCT Regulations and the PCT Administrative Instructions respectively.

INSTRUCTIONS CONCERNING AMENDMENTS UNDER ARTICLE 19

The applicant has, after having received the international search report, one opportunity to amend the claims of the international application. It should however be emphasized that, since all parts of the international application (claims, description and drawings) may be amended during the international preliminary examination procedure, there is usually no need to file amendments of the claims under Article 19 except where, e.g. the applicant wants the latter to be published for the purposes of provisional protection or has another reason for amending the claims before international pbulication. Furthermore, it should be emphasized that provisional protection is available in some States only.

What parts of the international application may be amended?

Under Article 19, only the claims may be amended.

During the international phase, the claims may also be amended (or further amended) under Article 34 before the International Preliminary Examining Authority. The description and drawings may only be amended under Article 34 before the International Examining Authority.

Upon entry into the national phase, all parts of the international application may be amended under Article 28 or, where applicable, Article 41.

When? Within 2 months from the date of transmittal of the international search report or 16 months from the priority date, whichever time limit expires later. It should be noted, however, that the amendments will be considered as having been received on time if they are received by the International Bureau after the expiration of the applicable time limit but before the completion of the technical preparations for international publication (Rule 46.1).

Where not to file the amendments?

The amendments may only be filed with the International Bureau and not with the receiving Office or the International Searching Authority (Rule 46.2).

Where a demand for international preliminary examination has been/is filed, see below.

How? Either by cancelling one or more entire claims, by adding one or more new claims or by amending the text of one or more of the claims as filed.

A replacement sheet must be submitted for each sheet of the claims which, on account of an amendment or amendments, differs from the sheet originally filed.

All the claims appearing on a replacement sheet must be numbered in Arabic numerals. Where a claim is cancelled, no renumbering of the other claims is required. In all cases where claims are renumbered, they must be renumbered consecutively (Administrative Instructions, Section 205(b)).

The amendments must be made in the language in which the international application is to be published.

What documents must/may accompany the amendments?

Letter (Section 205(b)):

The amendments must be submitted with a letter.

The letter will not be published with the international application and the amended claims. It should not be confused with the "Statement under Article 19(1)" (see below, under "Statement under Article 19(1)").

The letter must be in English or French, at the choice of the applicant. However, if the language of the international application is English, the letter must be in English; if the language of the international application is French, the letter must be in French.

NOTES TO FORM PCT/ISA/220 (continued)

The letter must indicate the differences between the claims as filed and the claims as amended. It must, in particular, indicate, in connection with each claim appearing in the international application (it being understood that identical indications concerning several claims may be grouped),whether

(i) the claim is unchanged;

(ii) the claim is cancelled;

(iii) the claim is new;

(iv) the claim replaces one or more claims as filed;

(v) the claim is the result of the division of a claim as filed.

The following examples illustrate the manner in which amendments must be explained in the accompanying letter:

1. [Where originally there were 48 claims and after amendment of some claims there are 51]:
"Claims 1 to 29, 31, 32, 34, 35, 37 to 48 replaced by amended claims bearing the same numbers; claims 30, 33 and 36 unchanged; new claims 49 to 51 added."

2. [Where originally there were 15 claims and after amendment of all claims there are 11]:
"Claims 1 to 15 replaced by amended claims 1 to 11."

3. [Where originally there were 14 claims and the amendments consist in cancelling some claims and in adding new claims]:
"Claims 1 to 6 and 14 unchanged; claims 7 to 13 cancelled; new claims 15, 16 and 17 added." or
"Claims 7 to 13 cancelled; new claims 15, 16 and 17 added; all other claims unchanged."

4. [Where various kinds of amendments are made]:
"Claims 1-10 unchanged; claims 11 to 13, 18 and 19 cancelled; claims 14, 15 and 16 replaced by amended claim 14; claim 17 subdivided into amended claims 15, 16 and 17; new claims 20 and 21 added."

"Statement under article 19(1)" (Rule 46.4)

The amendments may be accompanied by a statement explaining the amendments and indicating any impact that such amendments might have on the description and the drawings (which cannot be amended under Article 19(1)).

The statement will be published with the international application and the amended claims.

It must be in the language in which the international appplication is to be published.

It must be brief, not exceeding 500 words if in English or if translated into English.

It should not be confused with and does not replace the letter indicating the differences between the claims as filed and as amended. It must be filed on a separate sheet and must be identified as such by a heading, preferably by using the words "Statement under Article 19(1)."

It may not contain any disparaging comments on the international search report or the relevance of citations contained in that report. Reference to citations, relevant to a given claim, contained in the international search report may be made only in connection with an amendment of that claim.

Consequence if a demand for international preliminary examination has already been filed

If, at the time of filing any amendments under Article 19, a demand for international preliminary examination has already been submitted, the applicant must preferably, at the same time of filing the amendments with the International Bureau, also file a copy of such amendments with the International Preliminary Examining Authority (see Rule 62.2(a), first sentence).

Consequence with regard to translation of the international application for entry into the national phase

The applicant's attention is drawn to the fact that, where upon entry into the national phase, a translation of the claims as amended under Article 19 may have to be furnished to the designated/elected Offices, instead of, or in addition to, the translation of the claims as filed.

For further details on the requirements of each designated/elected Office, see Volume II of the PCT Applicant's Guide.

PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF ELECTION

(PCT Rule 61.2)

To:

Commissioner
US Department of Commerce
United States Patent and Trademark
Office, PCT
2011 South Clark Place Room
CP2/5C24
Arlington, VA 22202
ETATS-UNIS D'AMERIQUE

in its capacity as elected Office

Date of mailing (day/month/year) 09 July 2001 (09.07.01)	
International application No. PCT/EP00/09363	**Applicant's or agent's file reference** L10046 PCT
International filing date (day/month/year) 26 September 2000 (26.09.00)	**Priority date** (day/month/year) 01 October 1999 (01.10.99)
Applicant VALENCIA, Alfonso et al	

1. The designated Office is hereby notified of its election made:

[X] in the demand filed with the International Preliminary Examining Authority on:

26 April 2001 (26.04.01)

[ ] in a notice effecting later election filed with the International Bureau on:

2. The election [X] was

[ ] was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO **34, chemin des Colombettes** **1211 Geneva 20, Switzerland**	Authorized officer Zakaria EL KHODARY
Facsimile No.: (41-22) 740.14.35	Telephone No.: (41-22) 338.83.38

PCT

REQUEST

The undersigned requests that the present international application be processed according to the Patent Cooperation Treaty.

For [redacted]ing Office use only

International Application No.

International Filing Date

Name of receiving Office and "PCT International Application"

Applicant's or agent's file reference *(if desired) (12 characters maximum)* L10046 PCT

Box No. I TITLE OF INVENTION

PROCESS AND APPARATUS FOR IN SILICO TWO-HYBRID ANALYSIS

Box No. II APPLICANT

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

LION bioscience AG
Im Neuenheimer Feld 517
69120 Heidelberg
DE

☐ This person is also inventor.

Telephone No.

Facsimile No.

Teleprinter No.

State *(that is, country)* of nationality: DE

State *(that is, country)* of residence: DE

This person is applicant for the purposes of: ☐ all designated States ☒ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

Box No. III FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S)

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

Dr. VALENCIA, Alfonso
Azagador, 116.
Molino de la Hoz
Las Rozas
28230 Madrid, ES

This person is:

☐ applicant only

☒ applicant and inventor

☐ inventor only *(If this check-box is marked, do not fill in below.)*

State *(that is, country)* of nationality: ES

State *(that is, country)* of residence: ES

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☒ the United States of America only ☐ the States indicated in the Supplemental Box

☒ Further applicants and/or (further) inventors are indicated on a continuation sheet.

Box No. IV AGENT OR COMMON REPRESENTATIVE; OR ADDRESS FOR CORRESPONDENCE

The person identified below is hereby/has been appointed to act on behalf of the applicant(s) before the competent International Authorities as: ☒ agent ☐ common representative

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*

BOEHMERT & BOEHMERT
Dr. GODDAR, Heinz
Hollerallee 32
28209 Bremen
DE

Telephone No. 0421-34090

Facsimile No. 0421-3491768

Teleprinter No.

☐ **Address for correspondence:** Mark this check-box where no agent or common representative is/has been appointed and the space above is used instead to indicate a special address to which correspondence should be sent.

Sheet No. 2

Continuation of Box No. III FU[illegible]ER APPLICANT(S) AND/OR (FURTHER) I[illegible]NTOR(S)

If none of the following sub-boxes is used, this sheet should not be included in the request.

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

PAZOS, Florencio
Av. Monforte de Lemos, 83. bajo B.
28029 Madrid
ES

This person is:

- ☐ applicant only
- ☒ applicant and inventor
- ☐ inventor only *(If this check-box is marked, do not fill in below.)*

State *(that is, country)* of nationality: ES

State *(that is, country)* of residence: ES

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☒ the United States of America only ☐ the States indicated in the Supplemental Box

---

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

This person is:

- ☐ applicant only
- ☐ applicant and inventor
- ☐ inventor only *(If this check-box is marked, do not fill in below.)*

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

---

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

This person is:

- ☐ applicant only
- ☐ applicant and inventor
- ☐ inventor only *(If this check-box is marked, do not fill in below.)*

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

---

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

This person is:

- ☐ applicant only
- ☐ applicant and inventor
- ☐ inventor only *(If this check-box is marked, do not fill in below.)*

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

☐ Further applicants and/or (further) inventors are indicated on another continuation sheet.

Box No.V DESIGNATION OF STATES

The following designations are hereby made under Rule 4.9(a) *(mark the applicable check-boxes; at least one must be marked)*:

Regional Patent

- ☒ **AP** **ARIPO Patent:** **GH** Ghana, **GM** Gambia, **KE** Kenya, **LS** Lesotho, **MW** Malawi, **SD** Sudan, **SL** Sierra Leone, **SZ** Swaziland, **UG** Uganda, **ZW** Zimbabwe, and any other State which is a Contracting State of the Harare Protocol and of the PCT
- ☒ **EA** **Eurasian Patent:** **AM** Armenia, **AZ** Azerbaijan, **BY** Belarus, **KG** Kyrgyzstan, **KZ** Kazakhstan, **MD** Republic of Moldova, **RU** Russian Federation, **TJ** Tajikistan, **TM** Turkmenistan, and any other State which is a Contracting State of the Eurasian Patent Convention and of the PCT
- ☒ **EP** **European Patent:** **AT** Austria, **BE** Belgium, **CH and LI** Switzerland and Liechtenstein, **CY** Cyprus, **DE** Germany, **DK** Denmark, **ES** Spain, **FI** Finland, **FR** France, **GB** United Kingdom, **GR** Greece, **IE** Ireland, **IT** Italy, **LU** Luxembourg, **MC** Monaco, **NL** Netherlands, **PT** Portugal, **SE** Sweden, and any other State which is a Contracting State of the European Patent Convention and of the PCT
- ☒ **OA** **OAPI Patent:** **BF** Burkina Faso, **BJ** Benin, **CF** Central African Republic, **CG** Congo, **CI** Côte d'Ivoire, **CM** Cameroon, **GA** Gabon, **GN** Guinea, **GW** Guinea-Bissau, **ML** Mali, **MR** Mauritania, **NE** Niger, **SN** Senegal, **TD** Chad, **TG** Togo, and any other State which is a member State of OAPI and a Contracting State of the PCT *(if other kind of protection or treatment desired, specify on dotted line)* .....

National Patent *(if other kind of protection or treatment desired, specify on dotted line)*:

- ☒ **AE** United Arab Emirates
- ☒ **AL** Albania .....
- ☒ **AM** Armenia .....
- ☒ **AT** Austria .....
- ☒ **AU** Australia .....
- ☒ **AZ** Azerbaijan
- ☒ **BA** Bosnia and Herzegovina .....
- ☒ **BB** Barbados
- ☒ **BG** Bulgaria .....
- ☒ **BR** Brazil .....
- ☒ **BY** Belarus .....
- ☒ **CA** Canada
- ☒ **CH and LI** Switzerland and Liechtenstein
- ☒ **CN** China .....
- ☒ **CU** Cuba .....
- ☒ **CZ** Czech Republic .....
- ☒ **DE** Germany .....
- ☒ **DK** Denmark .....
- ☒ **EE** Estonia .....
- ☒ **ES** Spain .....
- ☒ **FI** Finland .....
- ☒ **GB** United Kingdom
- ☒ **GD** Grenada
- ☒ **GE** Georgia .....
- ☒ **GH** Ghana .....
- ☒ **GM** Gambia
- ☒ **HR** Croatia .....
- ☒ **HU** Hungary .....
- ☒ **ID** Indonesia
- ☒ **IL** Israel .....
- ☒ **IN** India .....
- ☒ **IS** Iceland
- ☒ **JP** Japan .....
- ☒ **KE** Kenya .....
- ☒ **KG** Kyrgyzstan .....
- ☒ **KP** Democratic People's Republic of Korea .....
- ☒ **KR** Republic of Korea .....
- ☒ **KZ** Kazakhstan .....
- ☒ **LC** Saint Lucia
- ☒ **LK** Sri Lanka
- ☒ **LR** Liberia
- ☒ **LS** Lesotho .....
- ☒ **LT** Lithuania
- ☒ **LU** Luxembourg
- ☒ **LV** Latvia
- ☒ **MD** Republic of Moldova .....
- ☒ **MG** Madagascar .....
- ☒ **MK** The former Yugoslav Republic of Macedonia .....
- ☒ **MN** Mongolia
- ☒ **MW** Malawi .....
- ☒ **MX** Mexico .....
- ☒ **NO** Norway
- ☒ **NZ** New Zealand .....
- ☒ **PL** Poland .....
- ☒ **PT** Portugal .....
- ☒ **RO** Romania
- ☒ **RU** Russian Federation .....
- ☒ **SD** Sudan
- ☒ **SE** Sweden
- ☒ **SG** Singapore
- ☒ **SI** Slovenia .....
- ☒ **SK** Slovakia .....
- ☒ **SL** Sierra Leone .....
- ☒ **TJ** Tajikistan .....
- ☒ **TM** Turkmenistan .....
- ☒ **TR** Turkey .....
- ☒ **TT** Trinidad and Tobago .....
- ☒ **UA** Ukraine .....
- ☒ **UG** Uganda .....
- ☒ **US** United States of America .....
- ☒ **UZ** Uzbekistan .....
- ☒ **VN** Viet Nam .....
- ☒ **YU** Yugoslavia .....
- ☒ **ZA** South Africa .....
- ☒ **ZW** Zimbabwe .....

Check-boxes reserved for designating States which have become party to the PCT after issuance of this sheet:

- ☐ .....
- ☐ .....

Precautionary Designation Statement: In addition to the designations made above, the applicant also makes under Rule 4.9(b) all other designations which would be permitted under the PCT except any designation(s) indicated in the Supplemental Box as being excluded from the scope of this statement. The applicant declares that those additional designations are subject to confirmation and that any designation which is not confirmed before the expiration of 15 months from the priority date is to be regarded as withdrawn by the applicant at the expiration of that time limit. *(Confirmation of a designation consists of the filing of a notice specifying that designation and the payment of the designation and confirmation fees. Confirmation must reach the receiving Office within the 15-month time limit.)*

See Notes to the request form

Box No. VI PRIORITY CLAIM ☐ Further priority claims are indicated in the Supplemental Box.

Filing date of earlier application (day/month/year)	Number of earlier application	Where earlier application is: national application: country	regional application:* regional Office	international application: receiving Office
item (1) 01.10.1999	99119515.7	EP		European Patent Office
item (2) 03.11.1999	99121794.4	EP		European Patent Office
item (3)				

☒ The receiving Office is requested to prepare and transmit to the International Bureau a certified copy of the earlier application(s) *(only if the earlier application was filed with the Office which for the purposes of the present international application is the receiving Office)* identified above as item(s): 1 and 2

* *Where the earlier application is an ARIPO application, it is mandatory to indicate in the Supplemental Box at least one country party to the Paris Convention for the Protection of Industrial Property for which that earlier application was filed (Rule 4.10(b)(ii)). See Supplemental Box.*

Box No. VII INTERNATIONAL SEARCHING AUTHORITY

Choice of International Searching Authority (ISA) *(if two or more International Searching Authorities are competent to carry out the international search, indicate the Authority chosen; the two-letter code may be used)* : ISA/

Request to use results of earlier search; reference to that search *(if an earlier search has been carried out by or requested from the International Searching Authority)*:

Date *(day/month/year)* Number Country *(or regional Office)*

Box No. VIII CHECK LIST; LANGUAGE OF FILING

This international application contains the following **number of sheets**:

request	:	4
description (excluding sequence listing part)	:	25
claims	:	8
abstract	:	1
drawings	:	5
sequence listing part of description	:	
Total number of sheets	:	43

This international application is **accompanied by** the item(s) marked below:

1. ☒ fee calculation sheet
2. ☐ separate signed power of attorney
3. ☐ copy of general power of attorney; reference number, if any:
4. ☐ statement explaining lack of signature
5. ☐ priority document(s) identified in Box No. VI as item(s):
6. ☐ translation of international application into (language):
7. ☐ separate indications concerning deposited microorganism or other biological material
8. ☐ nucleotide and/or amino acid sequence listing in computer readable form
9. ☒ other *(specify)*: copies of application papers (VI.1+2) for preparing priority documents

Figure of the drawings which should accompany the abstract: 1

Language of filing of the international application: English

Box No. IX SIGNATURE OF APPLICANT OR AGENT

Next to each signature, indicate the name of the person signing and the capacity in which the person signs (if such capacity is not obvious from reading the request).

Bremen, September 25, 2000

GODDAR, Heinz

For receiving Office use only

1. Date of actual receipt of the purported international application:	2. Drawings: ☐ received: ☐ not received:
3. Corrected date of actual receipt due to later but timely received papers or drawings completing the purported international application:	
4. Date of timely receipt of the required corrections under PCT Article 11(2):	
5. International Searching Authority (if two or more are competent): ISA/	6. ☐ Transmittal of search copy delayed until search fee is paid.

For International Bureau use only

Date of receipt of the record copy by the International Bureau:

This sheet is not part of and does not count as a sheet of the international application.

PCT

FEE CALCULATION SHEET

Annex to the Request

For receiving Office use only

International application No.

Date stamp of the receiving Office

Applicant's or agent's file reference: L10046 PCT

Applicant: LION bioscience AG

CALCULATION OF PRESCRIBED FEES

1. TRANSMITTAL FEE . . . . . EUR 102,-- T

2. SEARCH FEE . . . . . EUR 945,-- S

International search to be carried out by _____

(If two or more International Searching Authorities are competent in relation to the international application, indicate the name of the Authority which is chosen to carry out the international search.)

3. INTERNATIONAL FEE

Basic Fee

The international application contains 43 sheets.

first 30 sheets . . . . . EUR 409,-- b1

13 remaining sheets x EUR 9,-- additional amount = EUR 117,-- b2

Add amounts entered at b1 and b2 and enter total at B . . . . EUR 526,-- B

Designation Fees

The international application contains all designations.

8 number of designation fees payable *(maximum 10)* x EUR 88,-- amount of designation fee = EUR 704,-- D

Add amounts entered at B and D and enter total at I . . . . . EUR 1.230,-- I

(Applicants from certain States are entitled to a reduction of 75% of the international fee. Where the applicant is (or all applicants are) so entitled, the total to be entered at I is 25% of the sum of the amounts entered at B and D.)

4. FEE FOR PRIORITY DOCUMENT *(if applicable)* . . . . . EUR 60,-- P

5. TOTAL FEES PAYABLE . . . . . EUR 2.337,-- TOTAL

Add amounts entered at T, S, I and P, and enter total in the TOTAL box

☐ The designation fees are not paid at this time.

MODE OF PAYMENT

☐ authorization to charge deposit account (see below)	☐ bank draft	☐ coupons
☒ cheque	☐ cash	☐ other *(specify):*
☐ postal money order	☐ revenue stamps	

DEPOSIT ACCOUNT AUTHORIZATION *(this mode of payment may not be available at all receiving Offices)*

The RO/ _____

☐ is hereby authorized to charge the total fees indicated above to my deposit account.

☐ *(this check-box may be marked only if the conditions for deposit accounts of the receiving Office so permit)* is hereby authorized to charge any deficiency or credit any overpayment in the total fees indicated above to my deposit account.

☐ is hereby authorized to charge the fee for preparation and transmittal of the priority document to the International Bureau of WIPO to my deposit account.

Deposit Account No. _____ Date *(day/month/year)* _____ Signature _____

See Notes to the fee calculation sheet

PATENT COOPERATION TREATY

From the:
INTERNATIONAL PRELIMINARY EXAMINING AUTHORITY

To:
GODDAR, Heinz
BOEHMERT & BOEHMERT
Hollerallee 32
D-28209 Bremen
ALLEMAGNE

Boehmert & Boehmert
Bremen
Eing. 05. SEP 2001
Frist 04.11.01

PCT

WRITTEN OPINION

(PCT Rule 66)

Date of mailing *(day/month/year)*	04.09.2001
REPLY DUE	**within 2 month(s)** from the above date of mailing

Applicant's or agent's file reference	International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
L10046 PCT	PCT/EP00/09363	26/09/2000	01/10/1999

International Patent Classification (IPC) or both national classification and IPC
G06F19/00

Applicant
LION BIOSCIENCE AG

1. This written opinion is the **first** drawn up by this International Preliminary Examining Authority.

2. This opinion contains indications relating to the following items:

I	☒	Basis of the opinion
II	☐	Priority
III	☒	Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
IV	☒	Lack of unity of invention
V	☒	Reasoned statement under Rule 66.2(a)(ii) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
VI	☐	Certain document cited
VII	☐	Certain defects in the international application
VIII	☒	Certain observations on the international application

3. The applicant is hereby **invited to reply** to this opinion.

When? See the time limit indicated above. The applicant may, before the expiration of that time limit, request this Authority to grant an extension, see Rule 66.2(d).

How? By submitting a written reply, accompanied, where appropriate, by amendments, according to Rule 66.3. For the form and the language of the amendments, see Rules 66.8 and 66.9.

Also: For an additional opportunity to submit amendments, see Rule 66.4.
For the examiner's obligation to consider amendments and/or arguments, see Rule 66.4 bis.
For an informal communication with the examiner, see Rule 66.6.

If no reply is filed, the international preliminary examination report will be established on the basis of this opinion.

4. The final date by which the international preliminary examination report must be established according to Rule 69.2 is: 01/02/2002.

Name and mailing address of the international preliminary examining authority:
European Patent Office
D-80298 Munich
Tel. +49 89 2399 - 0 Tx: 523656 epmu d
Fax: +49 89 2399 - 4465

Authorized officer / Examiner
Seranski, P

Formalities officer (incl. extension of time limits)
Hingel, W
Telephone No. +49 89 2399 8717

I. Basis of the opinion

1. With regard to the **elements** of the international application (Replacement *sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this opinion as "originally filed"*):

Description, pages:

1-25 as originally filed

Claims, No.:

1-41 as originally filed

Drawings, sheets:

1/5-5/5 as originally filed

2. With regard to the **language**, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item.

These elements were available or furnished to this Authority in the following language: , which is:

- ☐ the language of a translation furnished for the purposes of the international search (under Rule 23.1(b)).
- ☐ the language of publication of the international application (under Rule 48.3(b)).
- ☐ the language of a translation furnished for the purposes of international preliminary examination (under Rule 55.2 and/or 55.3).

3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international preliminary examination was carried out on the basis of the sequence listing:

- ☐ contained in the international application in written form.
- ☐ filed together with the international application in computer readable form.
- ☐ furnished subsequently to this Authority in written form.
- ☐ furnished subsequently to this Authority in computer readable form.
- ☐ The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.
- ☐ The statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished.

4. The amendments have resulted in the cancellation of:

- ☐ the description, pages:
- ☐ the claims, Nos.:

☐ the drawings, sheets:

5. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2(c)):

   (Any replacement sheet containing such amendments must be referred to under item 1 and annexed to this report.)

6. Additional observations, if necessary:

III. Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

1. The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non-obvious), or to be industrially applicable have not been and will not be examined in respect of:

   ☐ the entire international application,

   ☒ claims Nos. 1-33, 35, 38-39,

because:

   ☒ the said international application, or the said claims Nos. 1-33, 35, 38-39 relate to the following subject matter which does not require an international preliminary examination (*specify*):
   see separate sheet

   ☐ the description, claims or drawings (*indicate particular elements below*) or said claims Nos. are so unclear that no meaningful opinion could be formed (*specify*):

   ☐ the claims, or said claims Nos. are so inadequately supported by the description that no meaningful opinion could be formed.

   ☐ no international search report has been established for the said claims Nos. .

2. A written opinion cannot be drawn due to the failure of the nucleotide and/or amino acid sequence listing to comply with the standard provided for in Annex C of the Administrative Instructions:

   ☐ the written form has not been furnished or does not comply with the standard.

   ☐ the computer readable form has not been furnished or does not comply with the standard.

IV. Lack of unity of invention

1. In response to the invitation (Form PCT/IPEA/405) to restrict or pay additional fees, the applicant has:

   ☐ restricted the claims.

   ☐ paid additional fees.

☐ paid additional fees under protest.

☐ neither restricted nor paid additional fees.

2. ☒ This Authority found that the requirement of unity of invention is not complied with for the following reasons and chose, according to Rule 68.1, not to invite the applicant to restrict or pay additional fees:
see separate sheet

3. Consequently, the following parts of the international application were the subject of international preliminary examination in establishing this opinion:

☐ all parts.

☒ the parts relating to claims Nos. 34, 36-37, 40-41.

V. Reasoned statement under Rule 66.2(a)(ii) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Claims	34, 36-37, 40-41
Inventive step (IS)	Claims	
Industrial applicability (IA)	Claims	

2. Citations and explanations
see separate sheet

VIII. Certain observations on the international application

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:
see separate sheet

1. Re Item III

Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

1.1 **Claims 1-33** relate to subject-matter considered by this Authority to be covered by the provisions of **Rule 67.1(iii) PCT**. Consequently, no opinion will be formulated with respect to the novelty, inventive step and industrial applicability of the subject-matter of these claims (Article 34(4)(a)(i) PCT).

1.2 **Claims 35 and 38-39** relate to subject-matter considered by this Authority to be covered by the provisions of **Rule 67.1(v) PCT**. Consequently, no opinion will be formulated with respect to the novelty, inventive step and industrial applicability of the subject-matter of these claims (Article 34(4)(a)(i) PCT).

2. Re Item IV

Lack of unity of invention

Lack of unity of the invention arises from the interpretation of the present set of claims, because the only technical feature that is common to all claims for which an examination can be carried out is that they are related by general methods used for the analysis of interaction between biomolecules. These methods are well known in the art. In consequence, claims related to physical entities like the pairs of interacting biomolecules of claim 34, computer readable media of claim 36-37 and computer systems of claim 40-41 have to be interpreted as single inventions.

3. Re Item V

Reasoned statement under Rule 66.2(a)(ii) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

3.1 For the assessment of novelty of present claims 34, 36-37 and 40-41 this authority is of the opinion, that the physical entities for which protection is sought are **differentiated** from physical entities of the prior art **only** by **non technical features.**

3.2 The Applicant tries to seek protection for the biomolecules of claim 34 by ways of

defining them by process features of a computer program. Claim 34, interpreted in favour of the applicant to be related to biological molecules and regardless of the problems arising from the clarity requirement (Art.6 PCT, see Point VIII), relates to features that are not of any technical nature and therefor do not contribute to a clear discrimination of the interacting biomolecules of claim 34 to any interacting biomolecules of the prior art, like for example DNA and polymerase or, more generally, any biological ligand binding to a receptor.

3.3 The same reasoning applies to the computer readable media of claims 36-37 and the computer systems of claims 40-41. In consequence, claims 34, 36-37 and 40-41 lack novelty as required by Art. 33(2) PCT.

4. Re Item VIII
Certain observations on the international application

Insufficiency of disclosure - Clarity (Art 83, 84 EPC)
4.1 The subject matter of claim 34 is not sufficiently disclosed as required by Art 5 PCT. Claim 34 refers to biomolecules identified by the methods of claim 2-31 without giving a true technical characterization. In addition, no such biomolecule is defined in the application. In consequence, the scope of the claim is ambiguous and vague and its subject matter is not sufficiently disclosed and supported as required by Art. 5 and 6 PCT

4.2 Claims 36-37 and 40-41, relating to computer readable media and computer systems, respectively, lack a clear technical characterization of the subject matter for which protection is sought, thus not fulfilling the requirements of Art.6 PCT.
In a further aspect the complete application is objected to for sufficiency of disclosure, as the information given in the description is merely based on scientific models and computer applications without giving a true technical teaching of how to perform the methods for which protection is being sought, i.e. the application neither teaches any program codes that can be used to perform the processes of the invention neither it enables a skilled person to come to the methods sought to be protected. In consequence, the application does not fulfil the requirements of Art.5 PCT.

5. If the Applicant thinks he could overcome above objections by filing a new set of

claims he is requested - in order to facilitate the examination of the conformity of the amended application with the requirements of Article 34(2)(b) PCT - to clearly identify the amendments carried out, no matter whether they concern amendments by addition, replacement or deletion, and to indicate the passages of the application as filed on which these amendments are based (see also Rule 66.8(a) PCT). It should be noted that this authority will disregard amendments for which no basis in the application as originally filed is indicated.

The demand must be filed directly with the competent International Preliminary Examining Authority; if two or more Authorities are competent, with the one chosen by the applicant. The full name or two-letter code of that Authority may be indicated by the applicant on the line below:

IPEA/ _____

PCT

CHAPTER II

DEMAND

under Article 31 of the Patent Cooperation Treaty:
The undersigned requests that the international application specified below be the subject of international preliminary examination according to the Patent Cooperation Treaty and hereby elects all eligible States (except where otherwise indicated).

For International Preliminary Examining Authority use only

Identification of IPEA	Date of receipt of DEMAND

Box No. I IDENTIFICATION OF THE INTERNATIONAL APPLICATION

Applicant's or agent's file reference: L10046PCT

International application No.	International filing date *(day/month/year)*	(Earliest) Priority date *(day/month/year)*
PCT/EP00/09363	26 September 2000 (26.09.00)	1 October 1999 (01.10.99)

Title of invention: PROCESS AND APPARATUS FOR IN SILICO TWO-HYBRID ANALYSIS

Box No. II APPLICANT(S)

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*

LION bioscience AG
Im Neuenheimer Feld 517
69120 Heidelberg
Germany

Telephone No.

Facsimile No.

Teleprinter No.

Applicant's registration No. with the Office

State *(that is, country)* of nationality: DE

State *(that is, country)* of residence: DE

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*

VALENCIA, Alfonso
Azagador, 116.
Molino de la Hoz
Las Rozas
28230 Madrid, Spain

State *(that is, country)* of nationality: ES

State *(that is, country)* of residence: ES

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*

PAZOS, Florencio
Av. Monforte de Lemos,
83. bajo B.
28029 Madrid
Spain

State *(that is, country)* of nationality: ES

State *(that is, country)* of residence: ES

☐ Further applicants are indicated on a continuation sheet.

International application No.
PCT/EP00/09363

Box No. III AGENT OR COMMON REPRESENTATIVE; OR ADDRESS FOR CORRESPONDENCE

The following person is ☒ agent ☐ common representative

and ☒ has been appointed earlier and represents the applicant(s) also for international preliminary examination.

☐ is hereby appointed and any earlier appointment of (an) agent(s)/common representative is hereby revoked.

☐ is hereby appointed, specifically for the procedure before the International Preliminary Examining Authority, in addition to the agent(s)/common representative appointed earlier.

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*

GODDAR, Heinz
BOEHMERT & BOEHMERT
Hollerallee 32
28209 Bremen
DE

Telephone No.
0421-34090

Facsimile No.
0421-3491768

Teleprinter No.

Agent's registration No. with the Office

☐ **Address for correspondence:** Mark this check-box where no agent or common representative is/has been appointed and the space above is used instead to indicate a special address to which correspondence should be sent.

Box No. IV BASIS FOR INTERNATIONAL PRELIMINARY EXAMINATION

Statement concerning amendments:*

1. The applicant wishes the international preliminary examination **to start on the basis of:**

☒ the international application as originally filed

the description ☐ as originally filed
☐ as amended under Article 34

the claims ☐ as originally filed
☐ as amended under Article 19 (together with any accompanying statement)
☐ as amended under Article 34

the drawings ☐ as originally filed
☐ as amended under Article 34

2. ☐ The applicant wishes any amendment to the claims under Article 19 to be considered as reversed.

3. ☐ The applicant wishes the start of the international preliminary examination **to be postponed** until the expiration of 20 months from the priority date unless the International Preliminary Examining Authority receives a copy of any amendments made under Article 19 or a notice from the applicant that he does not wish to make such amendments (Rule 69.1(d)). *(This check-box may be marked only where the time limit under Article 19 has not yet expired.)*

* Where no check-box is marked, international preliminary examination will start on the basis of the international application as originally filed or, where a copy of amendments to the claims under Article 19 and/or amendments of the international application under Article 34 are received by the International Preliminary Examining Authority before it has begun to draw up a written opinion or the international preliminary examination report, as so amended.

Language for the purposes of international preliminary examination: English

☒ which is the language in which the international application was filed.
☐ which is the language of a translation furnished for the purposes of international search.
☐ which is the language of publication of the international application.
☐ which is the language of the translation (to be) furnished for the purposes of international preliminary examination.

Box No. V ELECTION OF STATES

The applicant hereby **elects all eligible States** *(that is, all States which have been designated and which are bound by Chapter II of the PCT)*

excluding the following States which the applicant wishes **not to elect:**

Sheet No. 3

International application No.
PCT/EP00/09363

Box No. VI CHECK LIST

The demand is accompanied by the following elements, in the language referred to in Box No. IV, for the purposes of international preliminary examination:			For International Preliminary Examining Authority use only: received	not received
1. translation of international application	:	sheets	☐	☐
2. amendments under Article 34	:	sheets	☐	☐
3. copy (or, where required, translation) of amendments under Article 19	:	sheets	☐	☐
4. copy (or, where required, translation) of statement under Article 19	:	sheets	☐	☐
5. letter	:	sheets	☐	☐
6. other *(specify)*	:	sheets	☐	☐

The demand is also accompanied by the item(s) marked below:

1. ☒ fee calculation sheet
2. ☐ original separate power of attorney
3. ☐ original general power of attorney
4. ☐ copy of general power of attorney; reference number, if any:
5. ☐ statement explaining lack of signature
6. ☐ sequence listing in computer readable form
7. ☒ other *(specify)*: - cheque

Box No. VII SIGNATURE OF APPLICANT, AGENT OR COMMON REPRESENTATIVE

Next to each signature, indicate the name of the person signing and the capacity in which the person signs (if such capacity is not obvious from reading the demand).

For GODDAR, Heinz

SCHOHE, Stefan

For International Preliminary Examining Authority use only

1. Date of actual receipt of DEMAND:
2. Adjusted date of receipt of demand due to CORRECTIONS under Rule 60.1(b):
3. ☐ The date of receipt of the demand is AFTER the expiration of 19 months from the priority date and item 4 or 5, below, does not apply. ☐ The applicant has been informed accordingly.
4. ☐ The date of receipt of the demand is WITHIN the period of 19 months from the priority date as extended by virtue of Rule 80.5.
5. ☐ Although the date of receipt of the demand is after the expiration of 19 months from the priority date, the delay in arrival is EXCUSED pursuant to Rule 82.

For International Bureau use only

Demand received from IPEA on:

CHAPTER II

PCT

FEE CALCULATION SHEET

Annex to the Demand

International application No.	PCT/EP00/09363
Applicant's or agent's file reference	L10046PCT

For International Preliminary Examining Authority use only

Date stamp of the IPEA

Applicant: LION bioscience AG

CALCULATION OF PRESCRIBED FEES

1. Preliminary examination fee	EUR	1.533,00	P
2. Handling fee *(Applicants from certain States are entitled to a reduction of 75% of the handling fee. Where the applicant is (or all applicants are) so entitled, the amount to be entered at H is 25% of the handling fee.)*	EUR	147,00	H
3. Total of prescribed fees Add the amounts entered at P and H and enter total in the TOTAL box.....	EUR	1.680,00	TOTAL

MODE OF PAYMENT

- ☐ authorization to charge deposit account with the IPEA (see below)
- ☒ cheque
- ☐ postal money order
- ☐ bank draft
- ☐ cash
- ☐ revenue stamps
- ☐ coupons
- ☐ other *(specify):*

AUTHORIZATION TO CHARGE (OR CREDIT) DEPOSIT ACCOUNT

(This mode of payment may not be available at all IPEAs)

IPEA/ _____

☐ Authorization to charge the total fees indicated above.

Deposit Account No.: _____

☐ *(This check-box may be marked only if the conditions for deposit accounts of the IPEA so permit)* Authorization to charge any deficiency or credit any overpayment in the total fees indicated above.

Date: _____

Name: _____

Signature: _____

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 April 2001 (12.04.2001)

PCT

(10) International Publication Number
WO 01/26022 A1

(51) International Patent Classification[7]: G06F 19/00

(21) International Application Number: PCT/EP00/09363

(22) International Filing Date:
26 September 2000 (26.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
99119515.7 1 October 1999 (01.10.1999) EP
99121794.4 3 November 1999 (03.11.1999) EP

(71) Applicant *(for all designated States except US)*: **LION BIOSCIENCE AG** [DE/DE]; Im Neuenheimer Feld 517, 69120 Heidelberg (DE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **VALENCIA, Alfonso** [ES/ES]; Azagador, 116, Molino de la Hoz, Las Rozas, E-28230 Madrid (ES). **PAZOS, Florencio** [ES/ES]; Avenida Monforte de Lemos, 83 bajo B., E-28029 Madrid (ES).

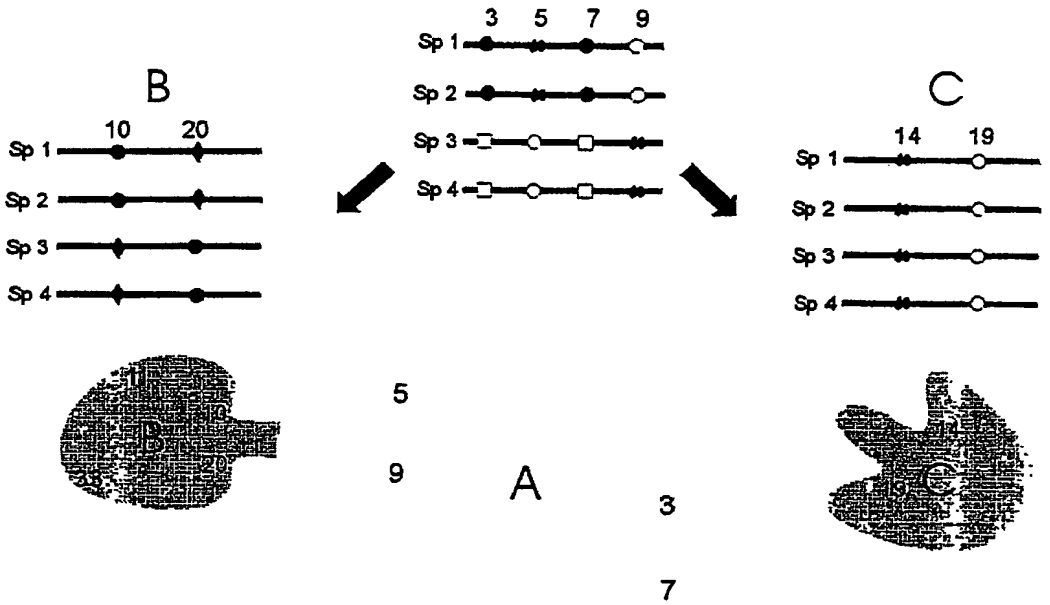
(74) Agent: **BOEHMERT & BOEHMERT**; Goddar, Heinz, Hollerallee 32, D-28209 Bremen (DE).

(81) Designated States *(national)*: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: PROCESS AND APPARATUS FOR IN SILICO TWO-HYBRID ANALYSIS



(57) Abstract: Process for the determination of interacting biomolecules wherein a) a first group is provided comprising sequences representing homologous biomolecules, b) at least one second group is provided comprising sequences representing homologous biomolecules, c) group correlation values between the sequences of the first group and the sequences of at least one second group are determined, and d) the probability of the interaction of the sequence represented biomolecules is determined on the basis of the group correlation values.

WO 01/26022 A1

Published:
— *With international search report.*
— *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Process and Apparatus for In Silico Two-Hybrid Analysis

The present invention relates to a process and a method for the determination of interacting biomolecules, use of such method, pairs of interacting biomolecules, data structure, computer readable medium, computer program, data base, computer system and device for simulating the interaction of biomolecules.

Background of the invention

Recently the advances in molecular biology have led to a vast amount of information at the genetic level of many different organisms, including man. In fact, there are attempts under way to determine the entire chromosome sequences of various different organisms, such as the human DNA sequence. Some have even been completed, such as the DNA sequence of the eubacteria *Escherichia coli*. A larger number of organisms with smaller genomes such as viruses or eu- and archaebacterial organisms have been sequenced and their predicted or assigned gene sequences lie in various public or private databases.

Advances in technology have led to what many people refer to as "reverse genetics", the analysis of the DNA sequence and the subsequent prediction and analysis of protein function.

Often, upon sequence determination, functional predictions are based on homology analysis of a particular protein to other proteins with well defined biochemical characteristics and function. In such case, the function is well known and the interactions between this protein and further proteins have been characterized biochemically in detail. Undoubtedly, one of the major factors of being able to attribute function to a protein is the knowledge of the interactions and its binding *in vivo* and *in vitro*.

The knowledge of the interactions and binding capabilities of proteins is not only of importance for function prediction but even more so for industrially relevant applications such as lead target interactions.

It is possible to differentiate between two types of interactions, strong interactions typical of structural proteins forming functional complexes and weak interactions more related with the transient coupling of proteins performing a function or control a system.

Examples of the first type of interaction and complexes thus formed are the ribosome, the proteosome or the spliceosome. All of them are big molecular complexes, with fixed components as well as interchangeable components. Smaller complexes, like the dimer of tubulin that forms microtubuli, or the histone components of the nucleosome are examples of stable structural interactions responsible for important cellular functions.

Examples of weak interactions forming, e.g., transient complexes, are the interaction between Elongation Factor Tu, a key initiator of protein translation, and other proteins (*i.e.* EF-Ts) regulating the transition between different states of EF-Tu (GTP bound active state and GDP bound inactive) or the interaction between DnaK, a molecular chaperon responsible of protein folding and transport in bacterial cells, and DnaJ which is a co-chaperon responsible for the regulation of activation of DnaK. Such transient complexes are difficult to study experimentally since their physical interaction is weaker than in the more stable structural complexes and their life time as a complex can be very short.

The emergence of new techniques in biology, namely the "Yeast Two Hybrid-Analysis" herein referred to as Y2H-analysis, a molecular approach which is described in US Patent 5,283,173, Mass-spectrometry applied to 2D gels and DNA-chips (for e.g. expression arrays), is leading to the experimental establishment of relations between two or more proteins of a given genome. These techniques are often in a developmental stage, require careful experimental set-up and considerable investments. All are subject to different types of experimental errors, and the determination of protein complexes and interactions have a considerable margin of error. Thus the determination of interacting protein pairs from a given set of possible pairings has been very tedious and often unsuccessful.

Mass spectroscopy combined with 2D gel electrophoresis can be used for identifying such proteins that change expression level or degree of post translational modification during varying biological process, *e.g.* heat shock. In this case a very considerable financial investment must be made and the performance of gel electrophoresis is often times tedious and time consuming.

Gene expression may also be analyzed using DNA-arrays. Here, the expression of gene products is monitored in different states of a cell. As a result it is sometimes possible to predict interacting proteins. However such expression analyses requires a great investment into fairly novel technology.

Until recently it was thus only possible to determine protein interactions by applying "wet" laboratory techniques, as outlined above. A number of computational techniques have however recently appeared trying to tackle this problem.

In those cases in which the three-dimensional structure of the two interacting proteins is known, docking techniques can be applied for modeling of the possible protein complex. These techniques are limited in their success, require intensive computational resources and, more importantly, can only be applied when the structures have been determined previously experimentally.

Another problem is the prediction of interacting regions between two proteins without prior knowledge of their three-dimensional structures. Pazos et al., (1997) have addressed this. They show that it is possible to predict which amino acids of the polypeptide chain are close in space, based on the information in the sequence that has accumulated over evolutionary time and can be retrieved from multiple sequence alignments of similar genes or proteins (*e.g.* multiple sequence alignment of the protein family). Still, this approach is focused on the prediction of the region of interaction between pairs of proteins known to interact.

The computational efforts to identify which proteins are likely to interact in the absence of experimental information has only been addressed very recently by Dandeker et al. (Conservation of gene order: a fingerprint of proteins that physically interact, (1998) Dandekar, T., Snel, B., Huynen, M. and Bork, P. TIBS 23:324-328) identifying a relation between the proximity in the genome of the genes in bacterial genomes and the probability of physical

interaction between their gene products. This approach is limited to a small set of genes whose proximity along the genome is conserved in many species.

Marcotte et al. (Marcotte et al., Science, vol. 285:751-753 (1999)) developed the idea of predicting protein-protein interaction for those multi-domain proteins that have different domain distributions in different genomes, *e.g.* a protein with two domains A and B from yeast may be represented in *E. coli* by two separate different proteins one containing domain A and the second containing domain B. The scope of this approach is limited to the few cases in which these types of molecular arrangements are met.

As outlined above there are "wet" laboratory techniques which enable the prediction of interacting sequences. These are often costly and time consuming. There are *in silico methods*, *i.e.* methods which can be adapted to run on computer devices for the prediction of interacting sites within a protein as well as interacting sites between two proteins known to interact but there is no method available for predicting the interaction biomolecules that takes only their sequences into account and that is generally applicable.

It was therefore an objective of the present invention to provide a process and/or a method which overcomes these limitations of the prior art.

It was also an objective of the present invention to provide for a process for identifying interacting biomolecules (i-2-hybrid process) that is reliable, cheap and avoids "wet" laboratory techniques.

It was therefore also an objective of the present invention to provide for a process that uses or requires only a primary sequence without the knowledge of a genome structure, the position of domains or other additional pieces of information.

It was a further objective of the present invention to provide for a process that may be applied to DNA, RNA and/or amino acid sequences.

It was further an objective of the present invention to provide for an apparatus for identifying interacting biomolecules.

It was further an objective of the present invention to create one or more databases containing information on such interacting sequence pairs by making use of the process of the present invention.

A further objective of the present invention is to provide a method for the determination of interacting biomolecules which comprises processing data. A still further object is to provide a data structure and a data base containing information and interacting sequence pairs.

Finally it is a further objective of the present application to provide a computer readable medium and a computer program product, respectively. Last but not least the objective of the present application is to provide a device for simulating the interaction of biomolecules represented by their sequences.

According to the invention this objective is solved by a process for the determination of interacting biomolecules and/or the simulation of the interaction of biomolecules wherein similar patterns of variation between two or more positions of at least two biomolecules are used.

According to the invention this objective is also solved by a process for the determination of interacting biomolecules which comprises the following steps:

a) a first group is provided comprising sequences representing homologous biomolecules,
b) at least one second group is provided comprising sequences representing homologous biomolecules,
c) group correlation values between the sequences of the first group and the sequences of at least one second group are determined, and
d) the probability of the interaction of the sequence represented biomolecules is determined on the basis of the group correlation values.

The objective is also solved by a method for the determination of interacting biomolecules which comprises processing data of at least a second set of data to output data

wherein each of the sets of data comprises independently and individually at least one or more elements,

wherein each of the elements represents the sequence of a biomolecule,

wherein the elements of a single set of data represent a group of homologous biomolecules

wherein the output data comprises at least one pair of elements with one part of the pair of elements comprising at least one element from the first set of data and the other part of the pair of elements comprising at least one element from the second set of data,

whereby

- a group correlation values data set is created comprising group correlation values which are determined between the sequences of the first set of data and at least the second set of data;

- an interaction probability data set is created by retrieving group correlation values from the group correlation values data set and determining the probability of interaction of the biomolecules based on the group correlation values; and

at least some of the elements from the first and at least the second set of data which have been used to create the group correlation values and the interaction probability therefrom form the output data.

Furthermore the objective is solved by the use of the inventive method for the simulation of biomolecule interaction.

The objective is furthermore solved by a data structure readable by a computer where said structure being generated by the inventive process or method.

The problem is also solved by a computer readable medium for embodying or storing therein data readable by a computer, said medium comprising one or more of the following:

a data structure generated by executing a process or a method;

Computer program code means which is adapted to cause a computer to execute the inventive process or method.

The problem is also solved by the database containing information on interesting sequence pairs generated by applying the inventive process or method.

The objective is also solved by a computer system comprising an execution environment for running their inventive process or inventive method.

The objective is furthermore solved by a device for simulating the interaction of biomolecules represented by the sequences which comprises

a loading device for making available the sets of data as specified in connection with the inventive method,

a processing device for performing the inventive method, and

an output device for receiving the output data generated by the processing device.

Finally the objective is also solved by pairs or complexes of interacting biomolecules determined in accordance with the inventive method or process.

In a preferred embodiment of the inventive process the probability of the interaction is calculated as predicted interaction value.

In a further preferred embodiment the interacting biomolecules are those with a positive predicted interaction value.

In a preferred embodiment of the inventive process any of the second group(s) is converted into the first group and the first group is converted into a second group and group correlation values between the sequences of this new first group and the sequences of any of the second group(s) which also comprises the former first group, are determined.

In a further embodiment of the inventive process site correlation values within each of the sequences within the first group and/or site correlation values within each of the sequences within the second group(s) are determined and said site correlation values are used for the calculation of the probability of interaction of the sequence represented biomolecules.

In a further more preferred embodiment of the inventive process the site correlation values are correlation values for substitutions within the sequences.

In a further embodiment of the inventive process each sequence of each of said groups is fused to each other to form fused sequences comprising at least one sequence of the first group and at least one sequence of any second group(s),

the correlation values within these fused sequences are determined, and

the correlation values are used as group correlation values for determining the probability of interaction.

In a further embodiment correlation values and preferably site correlation values are determined by

creating a position specific matrix containing the distances between pairs of sequences at that position whereby the distances are calculated by applying a standard distances matrix,

creating a combined matrix for two positions by calculating the covariation coefficient between corresponding entries of the position specific matrices, and

determining the correlation value for a pair of positions by averaging the correlation values of the combined matrix.

In a more preferred embodiment the standard distances matrix is the scoring matrix by McLachlan.

In an embodiment of the inventive method the probability of the interaction is calculated as predicted interaction value.

In a further embodiment the elements the predicted interaction value of which is positive, are interacting biomolecules.

In a still further embodiment any second set(s) of data is converted into the first set of data and the first set of data is converted into a second set of data, and

group correlation values are determined between the sequences of this new first set of data and the sequences of any of the second set(s).

In another embodiment of the inventive method site correlation values within each of the sequences within the first set of data and/or site correlation values within each of the sequences within the second set(s) of data are determined, and

said site correlation values form a set-specific site correlation value data set.

In a further embodiment of the inventive method the set-specific site correlation value data set is used to calculate the probability of interaction and/or to calculate the predicted interaction value of the sequence represented biomolecules.

In a further embodiment the site correlation values are correlation values for substitutions within the sequences.

In a still further embodiment a fused element set of data is generated by combining each element of the first set of data individually with each element of any of the second set(s) of data, and

attributing each fused element individually to the fused element set of data.

In a more preferred embodiment the correlation values are determined within the various positions of a single element of the fused element set of data, and

the correlation values are used as group correlation values for determining the probability of the interaction and/or predicted interaction value(s) of the biomolecules.

According to the inventive method the correlation values are determined in a preferred embodiment by

creating a position specific matrix containing the distances between pairs of sequences at that position whereby the distances are calculated by applying a standard distances matrix,

creating a combined matrix for two positions by calculating the covariation coefficient between corresponding entries of the position specific matrices or equivalent positions of the position specific matrices, and

determining the correlation value for a pair of positions by averaging the correlation values of the combined matrix.

More preferably the standard distances matrix is the scoring matrix by McLachlan.

In a further preferred embodiment of the inventive method the first set of data and/or the second set(s) of data are retrieved from a medium which is selected from the group comprising databanks, linked databanks, textual data and sets of data generated by an analytical instrument.

It is preferred that the set(s) of data comprise alligned sequences.

In another embodiment the output data are output control characters for a target medium.

In preferred embodiments of both the inventive method and process the sequences of the first group or second group(s) or first set of data or second set(s) of data are selected from the group comprising DNA sequences, RNA sequences and amino acid sequences.

In further preferred embodiments the number of sequences comprised in any of the groups or any of the sets of data is at least, preferably at least 11.

In another preferred embodiment of both the inventive method and process the sequences are homologous sequences.

In a more preferred embodiment the homologous sequences stem from different origins.

In an embodiment of both the inventive method and the inventive process the homologous sequences in the first set of data and in the second set of data stem from the same origin and/or the homologous sequence in the first group and in the second group stem from the same origin.

In further embodiments the homologous sequences are homologous genes.

In a more preferred embodiment the homologous genes are orthologs.

In a preferred embodiment of the inventive database the database is an organism/species specific database.

In an embodiment of use according to the present invention the interacting biomolecules are those with a positive predicted interaction value determined by a process or method according to the invention.

It has surprisingly been found that the yeast two hybrid system can be carried out in silico thus omitting the need for carrying out experiments in order to determine interacting biomolecules.

Herein homologues biomolecules represented by sequences means that the sequences of said biomolecules sequences are sequences that show sequence similarity. This sequence similarity may be high or low.

Herein, similar sequences are such sequences which are alignable when applying the CLUSTALW method (Higgins et al, see above) and such sequences that fall under the description of related sequences or derived sequences as found in

Doolittle, R. F. (1986). Of URFs and ORFs: a primer on how to analyze derived amino acid sequences. Mill Valley California: University Science Books.

or

McLachlan, A. D. (1971). Test for comparing related amino acid sequences. Mol. Biol. 61, 409-424.

Sequence homology may, but must not reflect *e.g.* that two or more sequences stem from a common origin or are otherwise related.

As outlined above the probability of the interaction of the biomolecules is determined on the basis of the correlation values. Such correlation values may be determined between elements, i.e. sequences, and more particularly distinct positions within these sequences, of the first group and the second group or, alternatively, the first set of data and the second set of data, respectively. These correlation values are referred to herein as group correlation values. These are to be understood as follows: A correlation value reflects similar patterns of variation of two or more positions within one sequence (referred to herein as site correlation values) or between the positions of sequences of two or more biomolecules (referred to herein as group correlation values). Similar patterns of variation are thought to be derived from simultaneous or concurrent events of sequence change along evolution. Such events are believed to reflect compensatory mutations. Herein, the use of the term "correlation value" does not imply the use of any particular algorithm or means of finding or determining such a value. Numerous means of determining correlation values may be applied.

The inventive process and method, respectively, make use of the correlation values, more particularly group correlation values. In a preferred embodiment of the process and method according to the invention this determination may be done as follows: the correlation values are normalized between 0 and 1 and divided in 10 levels. In principle however more or less than 10 levels may be applied.

Preferably, the percentage of correlated pairs, i.e. correlated pairs of sequences of the first and second group of sequences and sequences, i.e. elements, of the first and second set of data, of any correlation values, preferably the correlation values calculated for each alignment are calculated. Even more preferably, the sequences or elements of each group of set of data are aligned and the percentage of correlated pairs is calculated for the alignments of the individual sequences which are thus in some embodiment of the process/method multiple alignments and for the pairs between the two or more alignments. The correlation values are grouped in

levels of correlation to calculate the percentage of pairs correlated at each one of the levels. These percentages are used in a preferred embodiment subsequently in that part of the process/method where group correlation values are compared with the correlation values, also the site correlation values, preferably the correlation values of each one of the two alignments.

At each level of correlation the percentage of observations for the combined alignment (which corresponds to the fused sequence, see below) is divided by the sum of the values for the two individual alignments.

The result is multiplied by the value of correlation of the corresponding level. The final value for the prediction of probability of interaction between two biomolecules is obtained as the sum of the values calculated for the individual correlation levels. The probabilities are normalized by the average and standard deviation values (that is Z-score calculation). Average and standard deviations are calculated for the interaction of one biomolecule with all the other possible partners.

The process outlined above is shown graphically in Fig. 3.

In connection with the present invention the probability that two biomolecules interact is, in a particularly preferred embodiment, calculated based on a predicted interaction value. The calculation of this predicted interaction value is also described in connection with Fig. 3.

A positive predicted interaction value for two (or more) biomolecules means that there is a certain probability that said biomolecules will interact. The higher the predicted interaction value is, the higher the probability that said biomolecules will interact. In other words, the predicted interaction value is a measure for the probability that said biomolecules interact.

There are a multitude of ways of determining correlation values. These ways are known to the one skilled in the art and will be discussed later and are incorporated herein by reference. The process and method, respectively, according to the invention are not limited to one of these and may in fact make use of various different ways and methods, respectively, for determining the probability of the interaction of biomolecules based on correlation values.

Correlation values may be calculated as outlined in Göbel et al. (Göbel U, Sander C, Schneider R, Valencia A (1994). Correlated mutations and contact in proteins. Proteins 18, 309-317.) and modified to introduce a range correlation calculation. A position specific matrix is calculated for each position in the sequence. This position-specific matrix contains the distances between all sequence pairs at that position. Distances are defined by the scoring matrix of (McLachlan AD , 1971 J. Mol. Biol. 61, 409-424). Positions specific matrices are compared with a covariation coefficient formula that is applied to each of the corresponding values of the position specific matrices. The correlation between each pair of positions is calculated as the average of the covariation values. Fig. 2 outlines graphically the procedure described above.

In Altschuh et al. (Altschuh, D., Lesk, A.M., Bloomer, A.C., Klug, A. Correlation of co-ordinated amino acid substitutions with function in viruses related to tobacco mosaic virus. J.Mol.Biol.193: 693-707, 1987 and Altschuh, D., Verner, T., Berti, P., Moras, D., Nagai, K. Correlated amino acid changes in homologous protein families. Prot. Engin. 2: 193-199, 1988) correlation values are calculated as simple linear variation of identity patterns between subfamilies.

In Casari et al (Casari, G., Sander, C., Valencia, A., A method to predict functional residues in proteins. Nature Structural Biology. 2 (1995) 171-178) a principal component analysis method is applied to multiple sequence alignments to determine the correlation values between groups of positions.

In Lichtarge et al. (Lichtarge, O., Bourne, H. R., & Cohen, F. E. (1996), An Evolutionary Trace Method Defines Binding Surfaces Common to Protein Families. J. Mol. Biol. 257, 342-358) correlated positions are determined by careful manual analysis of phylogenetic trees in the search of positions clearly related with the main differences between tree branches.

Shindyalov et al. (Shindyalov, I. N., Kolchanov, N. A., & Sander, C. (1994). Can Three-Dimensional Contacts in Protein Structures be Predicted by Analysis of Correlated Mutations. Protein Eng. 7, 349-358) study the variation that accumulate simultaneously in different branches of phylogenetic trees. This method may also be applied for determining correlation values.

Taylor and Harrick (Taylor, W. R., & Harrick, K. (1994), Compensating Changes in Protein Multiple Sequence Alignments, Prot. Eng. 7, 342-348) describe a vector based method for the prediction of correlated mutations in multiple sequence alignments. The method takes into account physical properties and it is more related to the detection of simulatenous variation between different sub-families of proteins.

It is to be noted that any of the correlation values mentioned herein can be calculated in such a manner. Said techniques for calculating correlation values apply thus also to what is called group correlation values as well as to what is called site correlation values.

The methods known in the art for calculating correlation values take as a starting point the individual sequence (of a biomolecule). This applies also in connection with the site correlation values. In other words, the site correlation values are those which are calculated for various positions of a single sequence. This may be performed for any of the sequences in any of the groups or sets of data. The ratio behind this is to reduce the background for the calculation of the group correlation values (see also E.) in Fig. 3).

In a particular preferred embodiment of both the inventive process and the inventive method group correlation values are determined by actually forming one single sequence, which is called a fused sequence, of at least one sequence (or element) of the first group or of the first set of data and at least one sequence (or element) of the second group or of the second set of data. The created fused sequence is then used for the determination of the correlation values. By correlating a position of the fused sequence which stems originally from the first group or set of data, with a position of the fused sequence which stems originally from the second group or of set data, factually group correlation values can be determined.

Also because of this particular approach the determination of the site correlation value which is typically performed before the determination of the group correlation values and used to reduce the background of "wrong" or insignificant group correlation values, allows for the high accuracy of the inventive process and method.

In a preferred embodiment of the present invention the sequences in each group may be present in multiple sequence alignments.

Multiple sequence alignments used herein, refer to the alignment of DNA, RNA or amino acid sequences based on their sequence similarity. Such an alignment may be done manually or with the aid of a computer making use of an algorithm or method *e.g.* such algorithms or methods are *e.g.* the BLAST algorithm (Altschul, S.F., Gish, W. Miller, W., Myers, E.W., and Lipman, D.J., J. Mol. Biol. 215, 403-410 (1990) or an algorithm by Altschul (Altschul, S.F. (1993) "A protein alignment scoring system sensitive at all evolutionary distances." J. Mol. Evol. 36:290-300), the CLUSTALW method (Higgins, D. G., Bleasby, A. J., & Fuchs, R. (1992)), CLUSTAL V (Improved software for multiple sequence alignment. Comput. Appl. Biosci. 8, 189-191), or the MAXHOM method (Sander, C. & Schneider, R. (1991), Database of homology-derived structures and the structural meaning of sequence alignment, Proteins 9:56-68) but are not limited to these.

In a preferred embodiment of the present invention such alignments are generated using the Altschul algorithm (Altschul, S. F. (1993). A Protein Alignment Scoring System Sensitive at All Evolutionary Distances. J. Mol. Evol. 36, 290-300).

It should be noted that the above mentioned means of aligning sequences are examples or preferred embodiments. The process according to the invention can be realized using any method of sequence alignment.

In a preferred embodiment of the present invention additionally site correlation values for substitutions within the sequences within the first group or first set of data are determined, and additionally site correlation values for substitutions within the sequences within the further second group(s) or set(s) of data are determined. According to the invention such site correlation values may be used to determine the statistical significance of the group correlation values determined.

In a preferred embodiment of the process according to the invention the site correlation values within the groups are determined prior to the determination of the group correlation values.

A graphical representation of a possible embodiment of the process according to the present invention is depicted in Fig. 1.

The process and method according to the present invention also make use of one or more sequences. Such sequences may be RNA, DNA or amino acid sequences. They may be known, unknown i.e. generated de novo, publicly available or not. The sequences may be of natural origin or artificially generated, they may but must not represent genes or parts of genes. The sequences may have any given length.

Given one or more alignments to be used for the process according to the present invention it is desirable if each position in the alignment is coded by a distance matrix. It may be preferred if the distance matrix used is from McLachlan.

This position specific matrix contains all the residue-residue distances between all possible pairs of sequences at that position. These possible pairs may be pairs with both components of the pairs originating from a single group or set of data or with one component of the pairs originating from the first group or first set of data and the other component originating from the or any second group or set of data. In the event that a single fused sequence is created, pairs are also created. In this case the last residue of the fused sequence number 1 and the first residue of the fusion partner sequence number 2 are known thereby allowing for the above mentioned distance determination of the pairs of positions. Distances between amino acids may be defined by the scoring matrix of McLachlan (McLachlan, A. D. (1971). Test for comparing related amino acid sequences. J. Mol. Biol. 61, 409-424) or another scoring matrix.

The correlation value between each pair of positions is calculated as the average of the correlation for each corresponding bin of the position specific matrices. Corresponding bins contain the distance between the same two sequences in the two positions under comparison.

Bin as used herein means equivalent position in a matrix which is determined by a row and a column. The calculation implies comparing the corresponding positions of two "position specific matrices" calculating the covariation value for them and then averaging all the covariation values corresponding to each one of the different positions in the matrices (bins).

The DNA, RNA or amino acid sequences may stem from known or unknown organisms, may be created artificially, may represent sequences that are in parts from living or dead organisms and in other parts artificially created. The sequences may be newly determined using biochemical methods or may be taken from existing databases.

In a preferred embodiment the sequences represent genes of an organism. In another preferred embodiment the sequences represent the translated genes of an organism.

Most preferably the sequences are amino acid sequences and represent genes from an organism.

The process and method according to the present invention also make use of groups of sequences. In one embodiment of the invention such groups comprise at least 2 sequences. In an even more preferred embodiment of the invention such groups comprise at least 11 sequences. It should be noted that the more sequences are used the better the results achieved may become.

In a preferred embodiment of the present invention sequences in the groups stem from a multitude of different origins such as species, tissues or organisms representing a majority of sequence space and/or very distantly related species. Preferably one may want to align a number of sequences that are very similar that is about 50% similar or more, as well as sequences that are much less similar.

Herein, similar sequences are such sequences which are alignable when applying the CLUSTALW method (Higgins et al, see above) and such sequences that fall under the description of related sequences or derived sequences as found in

Doolittle, R. F. (1986). Of URFs and ORFs: a primer on how to analyze derived amino acid sequences. Mill Valley California: University Science Books.

or

McLachlan, A. D. (1971). Test for comparing related amino acid sequences. Mol. Biol. 61, 409-424.

Groups comprising very similar sequences as well as not very similar sequences, herein very similar are those sequences in which over 50 % of the residues are identical and not very

similar sequences are sequences in which 20 % or less of the residues are identical, are a preferred group for performing the process according to the invention.

In a preferred embodiment it is particularly advantageous if the sequences represented in the first group or set of data, the group containing homologous of the sequence representing biomolecules for which an interacting biomolecule is to be determined, stem from the same origin *i.e.* species, tissues or organisms as those sequences or elements in the second group or second set of data and *vice versa*. If, for example, origin means species the homologous biomolecules or their sequences may stem from different kinds of tissues of a single species such as liver or heart. If, as a further example, origin is to mean tissue, the homologous biomolecules or their sequences may stem from the same tissue of different species.

In a further preferred embodiment of the present invention it is desirable if the homologous sequences represent homologous genes. These genes may be represented by their amino acid sequence, their DNA sequence or their RNA sequence. A gene herein is to be understood as a DNA sequence that is transcribed into RNA *in vivo*, or a DNA or RNA sequence that encodes a polypeptide *in vivo*.

One can distinguish between orthologues sequences and paralogues sequences. Herein orthologues sequences are those which show close similarity between species and share a common evolutionary origin and paralogues sequences are those sequences which show close similarity within species, indicative of a close evolutionary relationship which may or may not have pre-dated speciation. The present invention may make use of both of these types of sequences. In a preferred embodiment of the present invention however, the process according to the invention makes use only of orthologues sequences.

The present invention can be performed manually or by using a computer. In a preferred embodiment of the present invention the data i.e. the sequences and/or the groups a) and/or b) of sequences are present in a computer readable form.

It is to be noted that what has been said in connection with the inventive process in principle also applies to the inventive method.

The inventive method may be used for the simulation of biomolecule interaction. This simulation is actually based on knowing which biomolecules may interact which each other. Insofar, the inventive method provides this prerequisite for the simulation. The simulation itself may then deduce from the sequence of the interacting biomolecules suitable representations such as three dimensional models to visualize the interaction.

The inventive device for simulating the interaction of biomolecules comprises, among others, a loading device and a processing device as well as an output device. The loading device may retrieve the sets of data required to perform the inventive method, e.g., from any kind of databank, analytical instrument, individual files, or textual information. This includes also retrieval of respective sets of data, i.e. sequences, from the internet. The processing device is then responsible for performing the inventive method and comprises preferably a computer. The processing device provides for output data which in turn are received from an output device for further handling of said output data. The output data may then be stored on any suitable medium, be printed out, written to a further document or be submitted to further processing.

Said device may also transfer the output data generated by the processing device making use of the inventive method as output control characters to, e.g. a further computer to perform the next step of the simulation of the interaction of the particular biomolecules where the fact that said biomolecules interact with each other, is carried out.

The present invention will further be illustrated by examples wherein

Fig. 1 illustrates the process by which a biomolecule A is analyzed with respect to interacting with biomolecules B and C;

Fig. 2 is a graphical representation of an embodiment of the proposed process for calculating correlated mutations;

Fig. 3 is a schematic representation of possible parts of the process according to the present invention;

Fig. 4 shows the results of the application of the process according to the invention; and

Fig. 5 shows the result of a determined interaction of biomolecules making use of the i-2-hybrid process.

Fig. 1 illustrates the process by which a biomolecule A is analyzed with respect to a possible interaction with biomolecules B and C in order to determine the most likely interaction partner. The process according to the present invention determines that the pattern of variation of two positions of the sequence biomolecule A (position 5 and 9) are similar and that they are at the same time similar to the patterns of variation of positions 10 and 20 of the sequence of protein B. No other positions are similar to them, for example none of the positions of the sequence of biomolecule C have patterns of variation similar to positions in protein A. Therefore the process according to the invention determines a proposed interaction between biomolecules A and B based on the possible interactions of positions 5 and 9 of sequence A with positions 10 and 20 of sequence B.

Fig. 2 is a graphical representation of an embodiment of the proposed process for calculating correlated mutations (Göbel et al., 1994). In A) a protein family is presented as a multiple sequence alignment (series of horizontal lines, where the numbers 1 to 3 represent different sequences (indices k, l run over proteins in the family) and the indices i and j run over positions in the alignment. Mutational behavior at each single position is summarized in a matrix B), including all the possible comparisons of the different sequences at that position. The position specific matrix C) is derived from B) according to a standard table of distances, *e.g.* McLachan (1971). In D) the covariation value is calculated for each one of the corresponding sequence pairs (k, l). Finally in E) the correlation value is calculated as the average of the covariation values of the two positions (i, j) and it carries information about the level of similarity of the mutational patterns of the two positions.

Fig. 3 shows a schematic representation of possible parts of the process according to the present invention. In A) sequences from different species (a, b, c, ...) are collected for two different biomolecules 1 and 2. The sequences are expected to correspond to the same species. In B) a virtual alignment is constructed concatenating the sequences of each one of the species for biomolecule 1 and 2. This concatenation leads to fusion sequences. The site correlation and group correlation values are calculated according to the procedure described above. In C) the correlation values are scaled (into correlation slots) between 0 and 1 and the frequency of

the different pairs of residues in the different correlation slots recorded. In the example the correlation values have been distributed in nine different correlation slots and the corresponding frequency distribution are represented for the two biomolecules (P_{11} and P22) and for the group correlation values (P_{12}).

D.) The distribution of site correlation values and the distribution of group correlation values are compared using the given formula. The correlated mutational behavior between the two biomolecules (C12) is calculated as the sum for the different correlation slots (i) of the ratio between distribution of values for the group correlation values (P12i) and the distribution of values for the site correlation values for the two biomolecules (P11i and P22i). This ratio is weighted with the value of the corresponding correlation slot (i) in a way that high values of correlation are given more importance.

E.) The predicted interaction value or predicted interaction score for the interaction between biomolecules 1 and 2 is calculated as a Z-score S12 of the C12 value relative to all interaction values for biomolecule 1. In the formula represented by C12 minus the mean of all interaction values for biomolecule 1 over the standard deviation of the group correlation values of the biomolecule 1 with all the other possible partners in the test set. The predicted interaction value is given in terms of standard deviation, positive values indicate a positive predicted interaction and the strength of the value the likelihood of the interaction.

Generally, the establishment of slots, more particularly correlation slots (i) refers to a process in which we take on the one hand the pairs of positions with the correlation value for the two possible alignments (site correlations), group them according to their value of correlation, count how many we have in each of the groups (for example, how many pairs in protein 1 are between 0.1 and 0.2 or correlation value) on the other hand we do the same for the pairs of position which form the alignment and the other forming the other (that is the pairs of the (group correlation). In this case we would be counting the number of pairs with one residue in protein 1 and another in protein 2, with a value of correlation between lets say 0.2 and 0.3. The process then is to compare the precentages of pairs at the different levels of correlation (the 0.2 to 0.3 level above) for the individual alignments site correlations with the percentage of pairs at the same level for the group correlation values. This is what is given in the formula.

Fig. 4 shows the result of the application of the process according to the invention (as exemplified in example 1) using sequences obtained from full genomes. The interaction probabilities are given in the Y-axis and proteins pairs are sorted according to these values (X-axis, logarithmic scale). The names of some of the proteins are indicated. The pairs of proteins known to interact are represented by filled symbols, those possible interactions corresponding in many cases to proteins that form part of complexes are given with open squares, most likely non-interacting proteins are represented by dots.

Fig. 5 shows the result of determined interaction of biomolecules making use of the i-2-hybrid (as specified in example 1) process depicted in a way reminiscent of the "wet" experimental yeast two hybrid system. Here, the diameter of the dots is proportional to the probability of interaction as determined using the process of the invention. In this case the minimal level of correlation entered in the analysis was of 0.4. The names of all the proteins used in the analysis are indicated. The empty squares correspond to those cases in which it was impossible to identify sequences from at least 11 species in common for those two proteins. The well known interacting proteins are highlighted with a dark square and the possibly interacting ones with a light-shaded square (*e.g.* different ribosomal proteins and elongation factors).

Examples:

Example 1 (Fig. 4 and 5):

The process according to the present invention was demonstrated by picking the right pair of interacting proteins in different sets of multiple sequence alignments. The multiple sequence alignments were generated using the ClustalW algorithm (Higgins, D. G., Bleasby, A. J., & Fuchs, R. (1992)).

A set composed of 53 proteins was analyzed. The sequences homologous to each one of them were collected from 14 different microbial genomes, that are completely sequenced and publicly available. The group correlation values were calculated for 244 pairs of proteins, that had at least 11 sequences from the same species in common.

In this set seven of the pairs of proteins with well documented interactions were among the ones with high predicted probability of interaction based on the group correlation values. And

additional set of ten pairs of proteins with high probability of interaction correspond to possibly interacting proteins as, for example, different ribosomal proteins. This high probability can be taken from or is expressed as a positive predicted interaction value herein.

Only one pair of proposed interacting biomolecules (SecD and SecF) has a relatively low value but has previously been described to represent interacting proteins. Interestingly, this probability is still better than any of the other probabilities calculated for these two proteins with the other alternative interaction partners in the set. The results of these experiments may be seen in figures 4 and 5.

Example 2:

Utilising the i-2-hybrid process we have constructed a database containing all the predicted interactions for *E. coli* proteins for which enough alignments were found in 20 complete bacterial genomes (4289 proteins as basic entries in the database). For this set of proteins it was possible to compute the interaction for 67238 pairs for which enough sequences of common species were detected. Each one of the entries is indexed and linked to other databases, in particular to Swissprot. The data base contains all the possible partners in each interaction and the reliability value of this interactions.

The quality of the predictions of interacting proteins in the database will benefit from continuous updates and from the continuous increase in the number of known sequences, in two ways. In the first place, the number of sequences that can be included in the alignments will raise the possibility of identifying interacting biomolecules and in the second place it will increase the reliability of the predictions since the basic methods often work better using alignments with many sequences.

Among the high scoring protein pairs determined by the i-2-hybrid process a number have previously been shown experimentally to interact as well, including membrane transporters of related compounds (G1787080-G1787369), transcriptions factors implicated in the control of related functions (G1787229-G1790863) or different subunits of an enzyme (G1787748-G2367325). As with other methods also false results may occur as the pair at position 10 (G1786981-G1790408) a transporter predicted to interact with a transcription factor/enzyme.

The database contains many interesting predictions of interaction, that in some cases could provide for a first clue with respect to the function of some proteins. For example among those comparisons resulting in positive predicted interaction values it is possible to find some proteins of known function like two transmembrane proteins (G1786670-G2367355) belonging to the (UPF0005) family, or another pair formed by a transcription factor implicated in the nitrate/nitrite response regulation that is predicted to interact with a protein of unknown function, that by homology looks similar to other transcription also implicated in nitrite/nitrate response regulation.

The database is organized to detect interactions for one given organism, in this case it is specific for *E. coli*, since the interactions are predicted in basis to the family alignments for a given organism.

The data contained in the newly generated database can for example be queried by protein names, gene names or accession numbers.

The features disclosed in the foregoing description, in the claims and/or in the accompanying drawings may, both separately and in any combination thereof, be material for realising the invention in diverse forms thereof.

Claims

1. Process for the determination of interacting biomolecules characterized in that similar patterns of variation between two or more positions of at least two biomolecules are used.

2. Process for the determination of interacting biomolecules, characterized in that

a) a first group is provided comprising sequences representing homologous biomolecules,

b) at least one second group is provided comprising sequences representing homologous biomolecules,

c) group correlation values between the sequences of the first group and the sequences of at least one second group are determined, and

d) the probability of the interaction of the sequence represented biomolecules is determined on the basis of the group correlation values.

3. Process according to claim 2, characterized in that the probability of the interaction is calculated as predicted interaction value.

4. Process according to claim 2 or 3, characterized in that the interacting biomolecules are those with a positive predicted interaction value.

5. Process according to any of claims 2 to 4, characterized in that any of the second group(s) is converted into the first group and the first group is converted into a second group and group correlation values between the sequences of this new first group and

the sequences of any of the second group(s) which also comprises the former first group, are determined.

6. Process according to any of claims 2 to 5, characterized in that
site correlation values within each of the sequences within the first group and/or site correlation values within each of the sequences within the second group(s) are determined and said site correlation values are used for the calculation of the probability of interaction and/or for the calculation of the predicted interaction value of the sequence represented biomolecules.

7. Process according to claim 6, characterized in that the site correlation values are correlation values for substitutions within the sequences

8. Process according to any of claims 2 to 7, characterized in that

each sequence of each of said groups is fused to each other to form fused sequences comprising at least one sequence of the first group and at least one sequence of any second group(s),

the correlation values within these fused sequences are determined, and

the correlation values are used as group correlation values for determining the predicted interaction value and/or the probability of interaction.

9. Process according to any of claims 2 to 8 characterized in that
correlation values are determined by

creating a position specific matrix containing the distances between pairs of sequences at that position whereby the distances are calculated by applying a standard distances matrix,

creating a combined matrix for two positions by calculating the covariation coefficient between equivalent positions of their position specific matrices, and

determining the correlation value for a pair of positions by averaging the correlation values of the combined matrix.

10. Process according to claim 9, characterized in that the standard distances matrix is the scoring matrix by McLachlan.

11. Method for the determination of interacting biomolecules which comprises processing data of at least a first set of data and at least a second set of data to output data

wherein each of the sets of data comprises independently and individually at least one or more elements,

wherein each of the elements represents the sequence of a biomolecule,

wherein the elements of a single set of data represent a group of homologous biomolecules,

wherein the output data comprises at least one pair of elements with one part of the pair of elements comprising at least one element from the first set of data and the other part of the pair of elements comprising at least one element from the second set of data,

characterised in that

- a group correlation values data set is created comprising group correlation values which are determined between the sequences of the first set of data and at least the second set of data;

- an interaction probability data set is created by retrieving group correlation values from the group correlation values data set and determining the probability of interaction of the biomolecules based on the group correlation values; and

at least some of the elements from the first and at least the second set of data which have been used to create the group correlation values and the interaction probability therefrom form the output data.

12. Method according to claim 11, characterized in that the probability of the interaction is calculated as predicted interaction value.

13. Method according to claim 11 or 12, characterized in that the elements the predicted interaction value of which is positive, are interacting biomolecules.

14. Method according to any of claims 11 to 13, characterized in that

any of second set(s) of data is converted into the first set of data and the first set of data is converted into a second set of data, and

group correlation values are determined between the sequences of this new first set of data and the sequences of any of the second set(s).

15. Method according to any of claims 11 to 14, characterized in that

site correlation values within each of the sequences within the first set of data and/or site correlation values within each of the sequences within the second set(s) of data are determined, and

said site correlation values form a set-specific site correlation value data set.

16. Method according to claim 15, characterized in that the set-specific site correlation value data set is used to calculate the probability of interaction of and/or to calculate the predicted interaction value of the sequence represented biomolecules.

17. Method according to claim 15 or 16, characterized in that the site correlation values are correlation values for substitutions within the sequences.

18. Method according to any of claims 11 to 17, characterized in that

a fused element set of data is generated by combining each element of the first set of data individually with each element of any of the second set(s) of data, and

attributing each fused element individually to the fused element set of data.

19. Method according to claim 18, characterized in that

the correlation values are determined within the various positions of a single element of the fused element set of data, and

the correlation values are used as group correlation values for determining the probability of the interaction of and/or predicted interaction value(s) of the biomolecules.

20. Method according to any one of claims 11 to 19, characterized in that the correlation values are determined by

creating a position specific matrix containing the distances between pairs of sequences at that position whereby the distances are calculated by applying a standard distances matrix,

creating a combined matrix for two positions by calculating the covariation coefficient between equivalent positions of their position specific matrices, and

determining the correlation value for a pair of positions by averaging the correlation values of the combined matrix.

21. Method according to claim 20, characterized in that the standard distances matrix is the scoring matrix by McLachlan.

22. Method according to any of claims 11 to 21, characterized in that the first set of data and/or second the second set(s) of data are retrieved from a medium which is selected from the group comprising databanks, linked databanks, textual data and sets of data generated by an analytical instrument.

23. Method according to any of claims 11 to 22, characterized in that the set(s) of data comprise alligned sequences.

24. Method according to any of claims 11 to 23, characterized in that the output data are output control characters for a target medium.

25. Method or process according to any of claims 2 to 24, characterized in that the sequences of the first group or second group(s) or first set of data or second set(s) of data are selected from the group comprising DNA sequences, RNA sequences and amino acid sequences.

26. Method or process according to any of claims 2 to 25, characterized in that the number of sequences comprised in any of the groups or any of the sets of data is at least , preferably at least 11.

27. Method or process according to any of claims 2 to 26, characterized in that the sequences are homologous sequences.

28. Method or process according to claim 27, characterized in that the homologous sequences stem from different origins.

29. Method or process according to claim 27, characterized in that the homologous sequences in the first set of data and in the second set of data stem from the same origin and/or the homologous sequence in the first group and in the second group stem from the same origin.

30. Method or process according to any of claims 27 to 29, characterized in that the homologous sequences are homologous genes.

31. Method or process according to claim 30, characterized in that the homologous genes are orthologs.

32. Use of the method according to any of claims 11 to 31 for the simulation of biomolecule interaction.

33. Use according to claim 32 wherein the interacting biomolecules are those with a positive predicted interaction value determined by a process or method according to any of the preceeding claims.

34. Pairs of interacting biomolecules determined according to a method or process according to any of the claims 2 to 31.

35. Data structure readable by a computer, said data structure being generated by a process or a method according to any of claims 2 to 31.

36. Computer readable medium for embodying or storing therein data readable by a computer, said medium comprising one or more of the following:

a data structure generated by executing a process or a method according to any of claims 2 to 31;

Computer program code means which is adapted to cause a computer to execute a process or method according to any one of claims 2 to 31.

37. Computer program product comprising the computer readable medium according to claim 36.

38. Database containing information on interacting sequence pairs generated by applying the process or method according to any of the claims 2 to 31.

39. Database according to claim 38, wherein the database is an organism/species specific database.

40. Computer system comprising an execution environment for running the process or method according to any of the claims 2 to 31.

41. Device for simulating the interaction of biomolecules represented by their sequences which comprises

a loading device for making available the sets of data according to any of the claims 11 to 31,

a processing device for performing the method according to any of the claims 11 to 31,

an output device for receiving the output data generated by the processing device.

Abstract

Process for the determination of interacting biomolecules,
wherein

a) a first group is provided comprising sequences representing homologous biomolecules,

b) at least one second group is provided comprising sequences representing homologous biomolecules,

c) group correlation values between the sequences of the first group and the sequences of at least one second group are determined, and

d) the probability of the interaction of the sequence represented biomolecules is determined on the basis of the group correlation values.

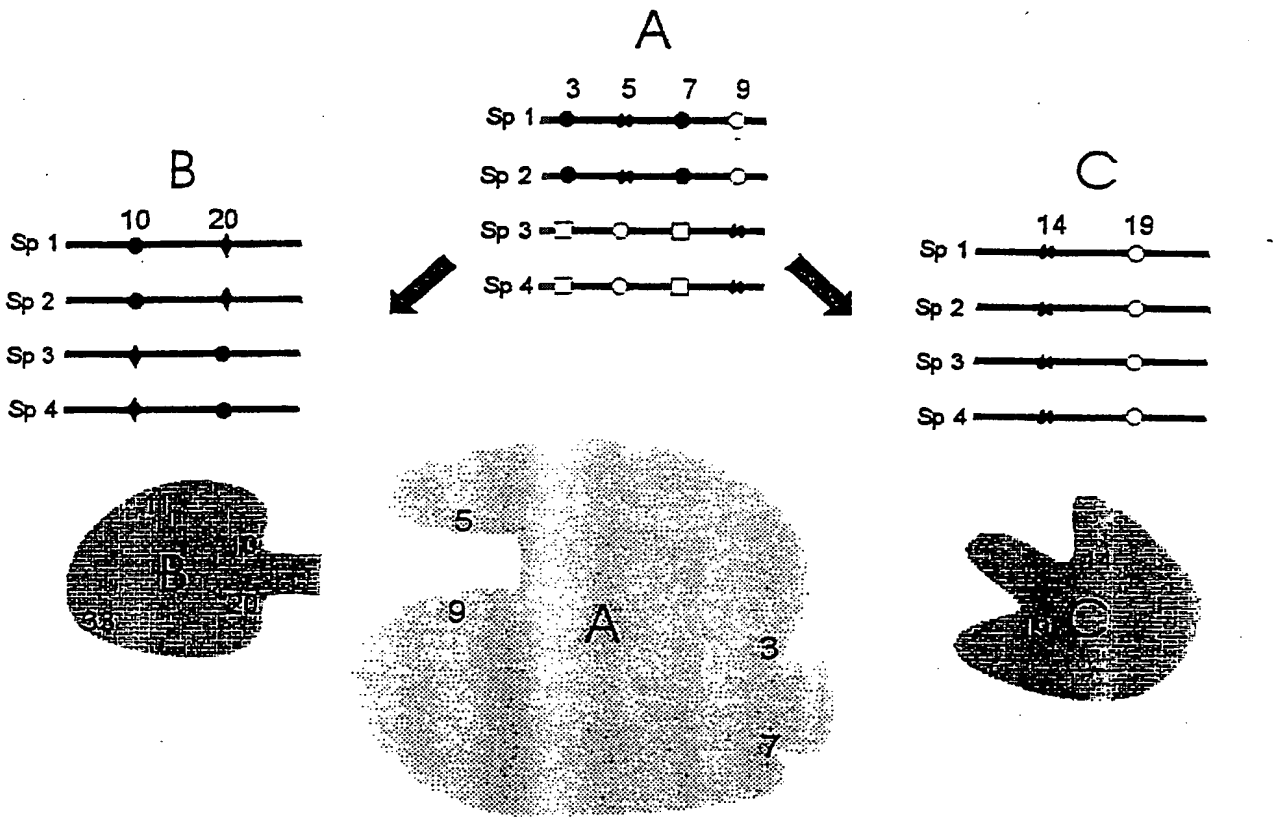
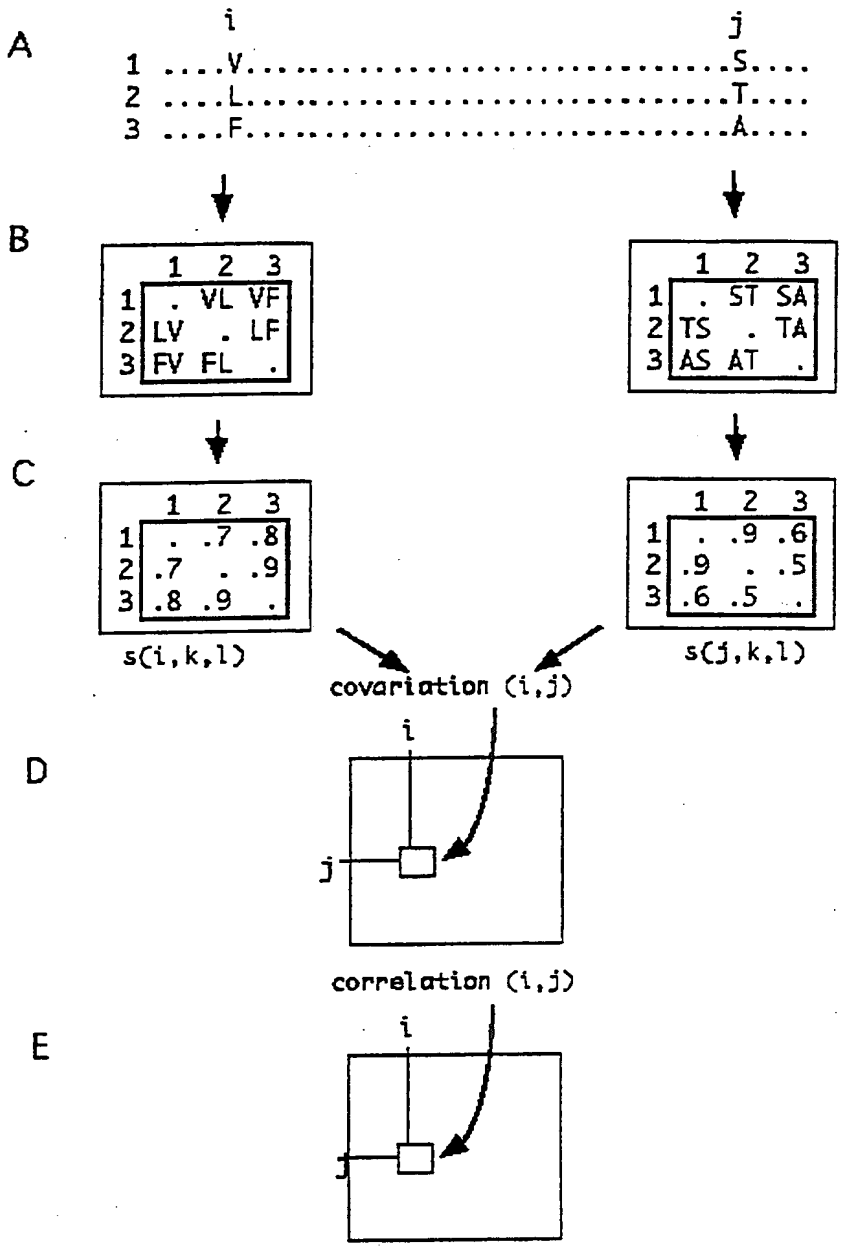
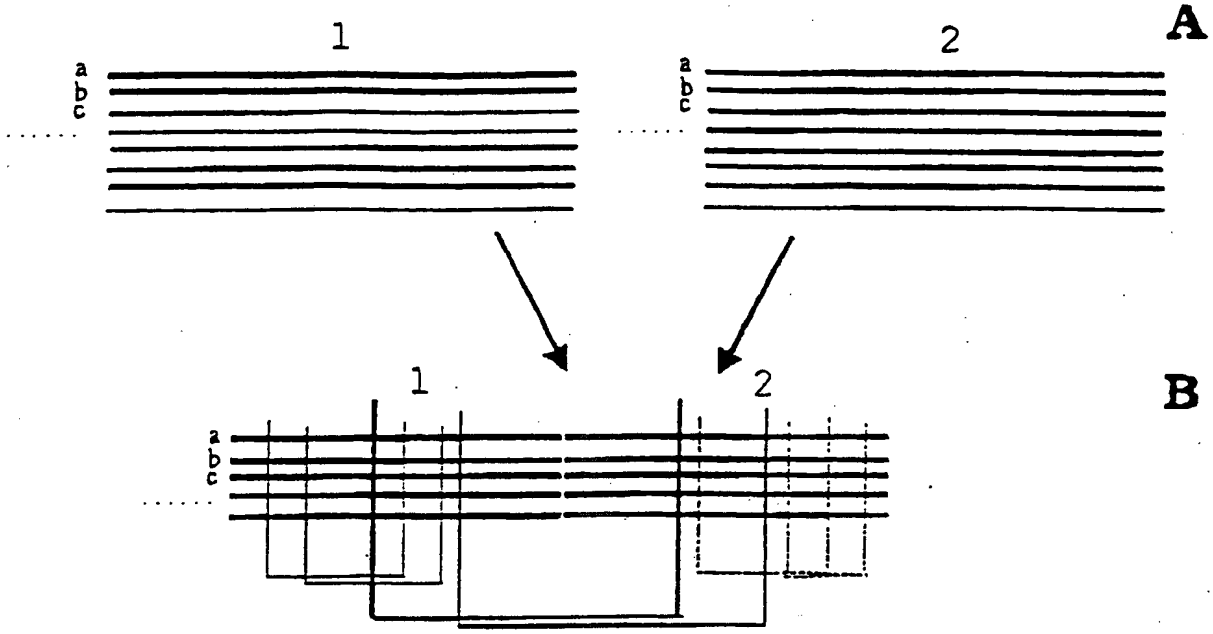
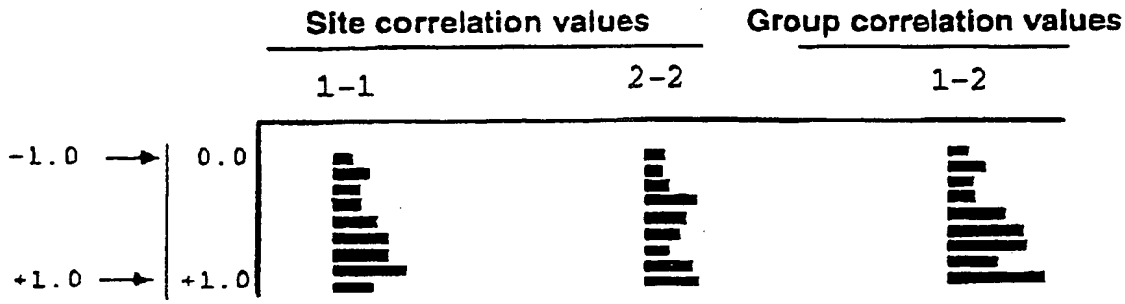

A
3 5 7 9
Sp 1
Sp 2
Sp 3
Sp 4
B
10 20
Sp 1
Sp 2
Sp 3
Sp 4
C
14 19
Sp 1
Sp 2
Sp 3
Sp 4
5
9
A
3
7

FIG. 1


A
i
j
1 ....V.....S....
2 ....L.....T....
3 ....F.....A....
B
1 2 3
1 . VL VF
2 LV . LF
3 FV FL .
1 2 3
1 . ST SA
2 TS . TA
3 AS AT .
C
1 2 3
1 . .7 .8
2 .7 . .9
3 .8 .9 .
s(i,k,l)
1 2 3
1 . .9 .6
2 .9 . .5
3 .6 .5 .
s(j,k,l)
covariation (i,j)
D
i
j
correlation (i,j)
E
i
j

F I G. 2





D

$$C12 = \sum_{i=0.0}^{1.0} \frac{P12i}{P11i + P22i} \, i$$

E

$$S12 = \frac{C12 - \overline{C1}}{\sigma 1}$$

FIG. 3

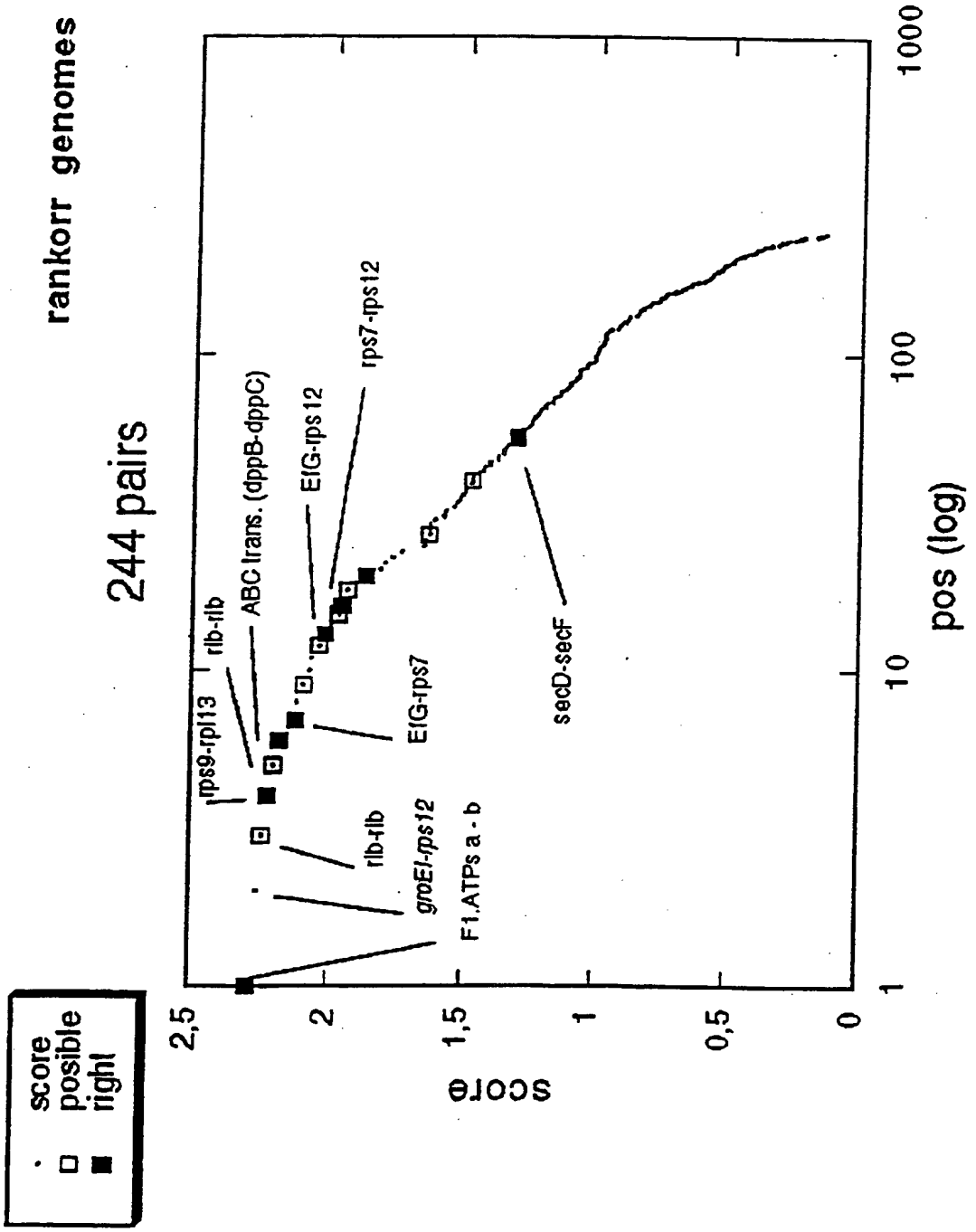

rankorr genomes
244 pairs
score
posible
right
rps9-rpl13
rlb-rlb
ABC trans. (dppB-dppC)
EfG-rps12
rps7-rps12
rlb-rlb
groEl-rps12
F1.ATPs a - b
EfG-rps7
secD-secF
score
pos (log)
2,5
2
1,5
1
0,5
0
1
10
100
1000

FIG. 4

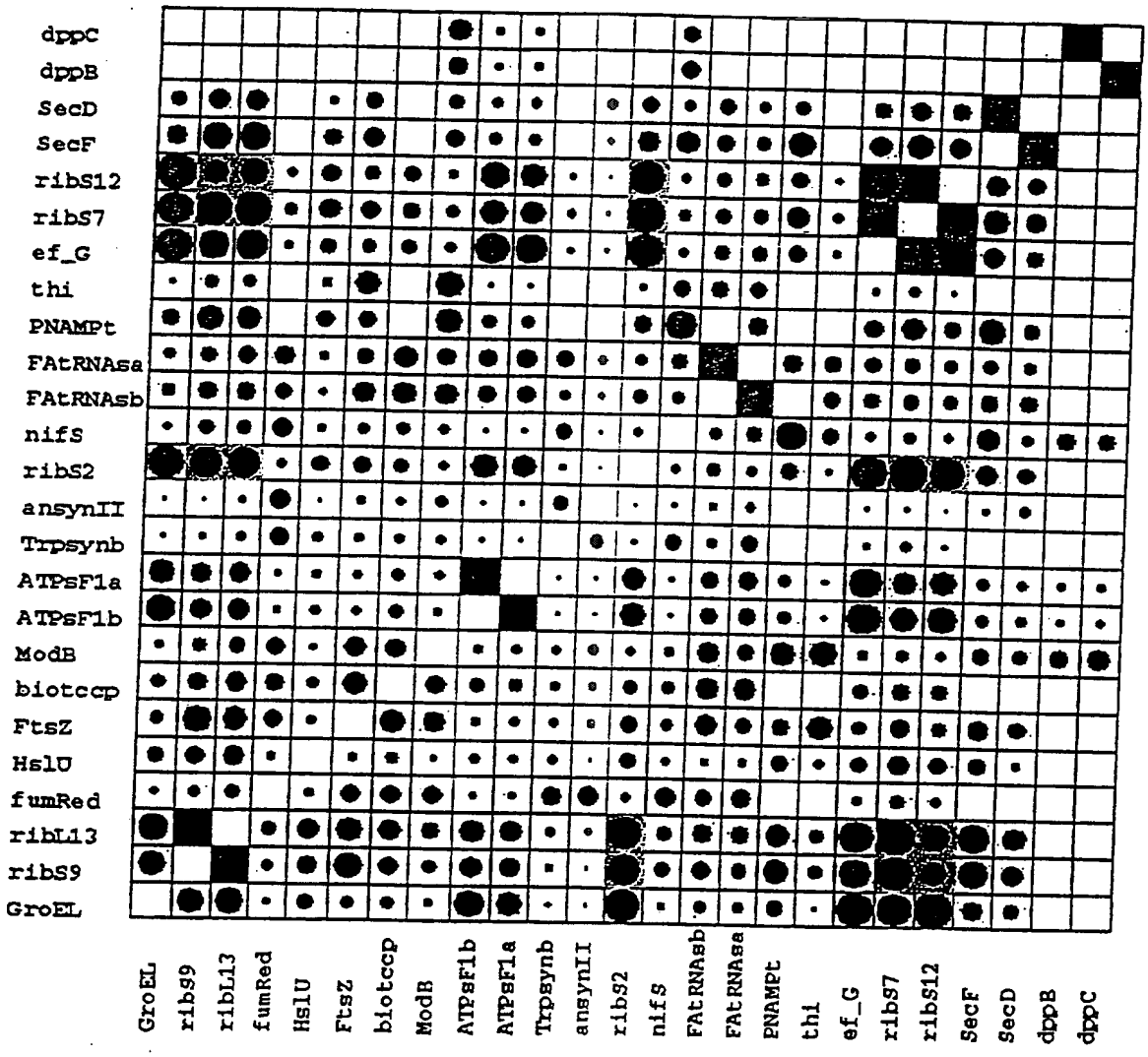

dppC
dppB
SecD
SecF
ribS12
ribS7
ef_G
thi
PNAMPt
FAtRNAsa
FAtRNAsb
nifS
ribS2
ansynII
Trpsynb
ATPsF1a
ATPsF1b
ModB
biotccp
FtsZ
HslU
fumRed
ribL13
ribS9
GroEL
GroEL
ribS9
ribL13
fumRed
HslU
FtsZ
biotccp
ModB
ATPsF1b
ATPsF1a
Trpsynb
ansynII
ribS2
nifS
FAtRNAsb
FAtRNAsa
PNAMPt
thi
ef_G
ribS7
ribS12
SecF
SecD
dppB
dppC

FIG. 5

12

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference	FOR FURTHER ACTION	See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)
L10046 PCT		
International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/EP00/09363	26/09/2000	01/10/1999
International Patent Classification (IPC) or national classification and IPC G06F19/00		
Applicant LION BIOSCIENCE AG		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 6 sheets, including this cover sheet.

   ☐ This report is also accompanied by ANNEXES, i.e. sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

   These annexes consist of a total of sheets.

3. This report contains indications relating to the following items:

   - I ☒ Basis of the report
   - II ☐ Priority
   - III ☒ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
   - IV ☒ Lack of unity of invention
   - V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations suporting such statement
   - VI ☐ Certain documents cited
   - VII ☐ Certain defects in the international application
   - VIII ☒ Certain observations on the international application

Date of submission of the demand	Date of completion of this report
26/04/2001	27.12.2001
Name and mailing address of the international preliminary examining authority: European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	Authorized officer Seranski, P Telephone No. +49 89 2399 7846

Form PCT/IPEA/409 (cover sheet) (January 1994)

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No. PCT/EP00/09363

I. Basis of the report

1. With regard to the **elements** of the international application *(Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report since they do not contain amendments (Rules 70.16 and 70.17))*:
Description, pages:

1-25 as originally filed

Claims, No.:

1-41 as originally filed

Drawings, sheets:

1/5-5/5 as originally filed

2. With regard to the **language**, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item.

These elements were available or furnished to this Authority in the following language: , which is:

- ☐ the language of a translation furnished for the purposes of the international search (under Rule 23.1(b)).
- ☐ the language of publication of the international application (under Rule 48.3(b)).
- ☐ the language of a translation furnished for the purposes of international preliminary examination (under Rule 55.2 and/or 55.3).

3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international preliminary examination was carried out on the basis of the sequence listing:

- ☐ contained in the international application in written form.
- ☐ filed together with the international application in computer readable form.
- ☐ furnished subsequently to this Authority in written form.
- ☐ furnished subsequently to this Authority in computer readable form.
- ☐ The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.
- ☐ The statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished.

4. The amendments have resulted in the cancellation of:

- ☐ the description, pages:
- ☐ the claims, Nos.:

☐ the drawings, sheets:

5. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2(c)):

(Any replacement sheet containing such amendments must be referred to under item 1 and annexed to this report.)

6. Additional observations, if necessary:

III. Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

1. The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non-obvious), or to be industrially applicable have not been examined in respect of:

☐ the entire international application.

☒ claims Nos. 1-33, 35, 38-39.

because:

☒ the said international application, or the said claims Nos. 1-33, 35, 38-39 relate to the following subject matter which does not require an international preliminary examination (*specify*):
see separate sheet

☐ the description, claims or drawings (*indicate particular elements below*) or said claims Nos. are so unclear that no meaningful opinion could be formed (*specify*):

☐ the claims, or said claims Nos. are so inadequately supported by the description that no meaningful opinion could be formed.

☐ no international search report has been established for the said claims Nos. .

2. A meaningful international preliminary examination cannot be carried out due to the failure of the nucleotide and/or amino acid sequence listing to comply with the standard provided for in Annex C of the Administrative Instructions:

☐ the written form has not been furnished or does not comply with the standard.

☐ the computer readable form has not been furnished or does not comply with the standard.

IV. Lack of unity of invention

1. In response to the invitation to restrict or pay additional fees the applicant has:

☐ restricted the claims.

☐ paid additional fees.

☐ paid additional fees under protest.

☐ neither restricted nor paid additional fees.

2. ☒ This Authority found that the requirement of unity of invention is not complied and chose, according to Rule 68.1, not to invite the applicant to restrict or pay additional fees.

3. This Authority considers that the requirement of unity of invention in accordance with Rules 13.1, 13.2 and 13.3 is

☐ complied with.

☒ not complied with for the following reasons:
see separate sheet

4. Consequently, the following parts of the international application were the subject of international preliminary examination in establishing this report:

☐ all parts.

☒ the parts relating to claims Nos. 34, 36-37, 40-41.

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims	
	No:	Claims	34, 36-37, 40-41
Inventive step (IS)	Yes:	Claims	
	No:	Claims	34, 36-37, 40-41
Industrial applicability (IA)	Yes:	Claims	34, 36-37, 40-41
	No:	Claims	

2. Citations and explanations
see separate sheet

VIII. Certain observations on the international application

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:
see separate sheet

1. Re Item III

Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

1.1 **Claims 1-33** relate to subject-matter considered by this Authority to be covered by the provisions of **Rule 67.1(iii) PCT**. Consequently, no opinion will be formulated with respect to the novelty, inventive step and industrial applicability of the subject-matter of these claims (Article 34(4)(a)(i) PCT).

1.2 **Claims 35 and 38-39** relate to subject-matter considered by this Authority to be covered by the provisions of **Rule 67.1(v) PCT**. Consequently, no opinion will be formulated with respect to the novelty, inventive step and industrial applicability of the subject-matter of these claims (Article 34(4)(a)(i) PCT).

2. Re Item IV

Lack of unity of invention

Lack of unity of the invention arises from the interpretation of the present set of claims, because the only technical feature that is common to all claims for which an examination can be carried out is that they are related by general methods used for the analysis of interaction between biomolecules. These methods are well known in the art. In consequence, claims related to physical entities like the pairs of interacting biomolecules of claim 34, computer readable media of claim 36-37 and computer systems of claim 40-41 have to be interpreted as single inventions.

3. Re Item V

Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

3.1 For the assessment of novelty of present claims 34, 36-37 and 40-41 this authority is of the opinion, that the physical entities for which protection is sought are **differentiated** from physical entities of the prior art **only** by **non technical features.**

3.2 The Applicant tries to seek protection for the biomolecules of claim 34 by ways of

defining them by process features of a computer program. Claim 34, interpreted in favour of the applicant to be related to biological molecules and regardless of the problems arising from the clarity requirement (Art.6 PCT, see Point VIII), relates to features that are not of any technical nature and therefor do not contribute to a clear discrimination of the interacting biomolecules of claim 34 to any interacting biomolecules of the prior art, like for example DNA and polymerase or, more generally, any biological ligand binding to a receptor.

3.3 The same reasoning applies to the computer readable media of claims 36-37 and the computer systems of claims 40-41. In consequence, claims 34, 36-37 and 40-41 lack novelty as required by Art. 33(2) PCT.

4. Re Item VIII

Certain observations on the international application

Insufficiency of disclosure - Clarity (Art 83, 84 EPC)

4.1 The subject matter of claim 34 is not sufficiently disclosed as required by Art 5 PCT. Claim 34 refers to biomolecules identified by the methods of claim 2-31 without giving a true technical characterization. In addition, no such biomolecule is defined in the application. In consequence, the scope of the claim is ambiguous and vague and its subject matter is not sufficiently disclosed and supported as required by Art. 5 and 6 PCT

4.2 Claims 36-37 and 40-41, relating to computer readable media and computer systems, respectively, lack a clear technical characterization of the subject matter for which protection is sought, thus not fulfilling the requirements of Art.6 PCT.
In a further aspect the complete application is objected to for sufficiency of disclosure, as the information given in the description is merely based on scientific models and computer applications without giving a true technical teaching of how to perform the methods for which protection is being sought, i.e. the application neither teaches any program codes that can be used to perform the processes of the invention neither it enables a skilled person to come to the methods sought to be protected. In consequence, the application does not fulfil the requirements of Art.5 PCT.

BOEHMERT & BOEHMERT
ANWALTSSOZIETÄT

Boehmert & Boehmert • P.O.B. 15 03 08 • D-80043 München

Europäisches Patentamt
Erhardtstr. 27
80298 München

Ihr Zeichen Your ref.	Ihr Schreiben Your letter of	Unser Zeichen Our ref.	**München,**
PCT/EP00/09363		L10046PCT	November 5, 2001

International Patent Application No. PCT/EP00/09363
"Process and Apparatus for in Silico Two-Hybrid Analysis"
LION bioscience AG

In response to the written opinion of the IPEA dated 04.09.2001

A. Re Item III:

1. Objections based on Art. 34 (4) a), Rule 67.1 (iii) PCT, Claims 1 to 33:

The examining authority states that claims 1-33 relate to subject-matter considered to be covered by the provisions of **Art. 34 (4) a) and Rule 67. (iii) PCT.**

The invention underlying the present international patent application relates to a process for the determination of interacting biomolecules, wherein similar patterns of variation between two or more positions of at least two biomolecules are used (claim 1).

Pettenkoferstraße 20-22 · D-80336 München · P.O.B. 15 03 08 · D-80043 München · Telephon +49-89-559680 · Telefax +49-89-347010
MÜNCHEN - BREMEN - BERLIN - DÜSSELDORF - FRANKFURT- BIELEFELD - POTSDAM - BRANDENBURG - KIEL - PADERBORN - HÖHENKIRCHEN - ALICANTE
http://www.boehmert.de e-mail: postmaster@boehmert.de

In particular, the application relates to a process for the determination of interacting biomolecules, comprising

Providing a first group a) of biomolecules represented by their homologues sequences,

providing one or more further groups b) comprising biomolecules represented by their homologues sequences,

determining group correlation values between the sequences of group a) and the sequences of at least one further group b) and

determining the probability of the interaction of biomolecules on the basis of the group correlation values,

and selecting interacting biomolecules from the group comprising those with high probabilities (claim 2).

In claim 11, the invention further relates to a method for the determination of interacting biomolecules which comprises processing data of at least a first set of data and at least a second set of data to output data wherein, each of the sets of data comprises independently and individually at least one or more elements, wherein each of the elements represents the sequence of a biomolecule, wherein the elements of a single set of data represent a group of homologous biomolecules, wherein the output data comprises at least one pair of elements with one part of the pair of elements comprising at least one element from the first set of data and the other part of the pair of elements comprising at least one element from the second set of data, characterised in that a group correlation values data set is created comprising group correlation values which are determined between the sequences of the first set of data and at least the second set of data, an interaction probability data set is created by retrieving group correlation values from the group correlation values data set and determining the probability of interaction of the biomolecules based on the group correlation values; and at least some of the elements from the first and at least the second set of data which have been used to create the group correlation values and the interaction probability therefrom form the output data.

Thus, the methods as claimed in claims 1-33 relate to the determination of interacting biomolecules which comprises processing physical data with respect to the biomolecules that are used in said processes. According to the opinion of, for example, the Boards of Appeal of the EPO, physical data can be present in the form of parameters, i. e. in this case parameters of the biomolecules. Furthermore, this processing of physical data is regarded as technical in relation to computer-implemented inventions. It is apparent from the description that the invention is preferably performed on a computer (see for example page 19, 5th paragraph), i. e. computer-implemented. According to the opinion of the applicant, the criteria of the decision of the Board of Appeal of the EPO, T1173/97 are applicable.

Again, for the sake of clarity applicant submits that exact wording of Rule 67.1 reads: "No International Preliminary Examining Authority shall be required to carry out an international preliminary examination on an international application if, and to the extent to which, its subject matter is any of the following: (...)

(iii) schemes, rules or methods of doing business, performing purely mental acts or playing games, (...).

As outlined above the claims 1-33 relate to methods for the determination of interacting biomolecules which comprise processing physical data. Claims 1 to 33 further cite various steps wherein physical data sets are handled, analysed and or further processed.

Based on the wording of claims 11 to 33, applicant submits that the claimed method is neither a scheme, rule or method of doing business, the performing of purely mental acts nor is it a method for game playing.

Also, in light of the objective problem solved by the invention, which can be formulated as "the provision of a method for identifying interacting proteins" and which is similar if not identical to that of the y-2-h analysis *in-vitro* method, the invention as claimed in claims 1 - 33 can neither be considered to be a) a scheme, b) a rule, c) a method of doing business, d) the performance of a purely mental act, or e) the playing of a game. This in particular holds true for claim 11 of the present international application.

The steps performed by the invention as claimed, wherein data sets are handled, and processed, are not of administrative, actuarial and/or financial character which would be typical for a business method.

Thus, the subject matter as claimed, when applying the "whole claim approach", provides a contribution to the art in a field NOT excluded from patentability under PCT Art. 34 (4) a).

The invention furthermore claims a device which "handles" data sets and performs the method as described. This device, of course, is technical.

Taken together, the applicant submits that the Preliminary Examining Authority is incorrect when considering the subject matter of claims 1-33 to be covered by Rule 67.1 (iii). A reconsideration of this ruling and consequently the formulation of an opinion based on novelty, industrial applicability and inventive step considerations is respectfully requested.

2. Objections based on Art. 34 (4) a), Rule 67.1 (v) PCT, Claims 35 and 38 to 39:

Claim 35:

Claim 35 relates to a physical data structure readable by a computer, said data structure being generated by a process or a method according to the invention. The data structure represents a similar

result as that which is achieved when performing the *in-vitro* "wet" y-2-h-analysis, namely the probability of the *in-vivo* interaction of two biomolecules, a parameter with respect to the two proteins examined. In contrast to the y-2-h-analysis, the present invention enables the creation of large databases comprising numerous probabilities of such interactions.

The "data structure readable by a computer" is claimed by way of a "product-by-process-claim" as the application does not enable the adequate description of the data structure other than by way of its creation.

The applicant submits, that the examining authority is incorrect when considering the subject matter of claim 35 to be covered by Rule 67.1 (v). A reconsideration of this ruling and consequently the formulation of an opinion based on novelty and inventive step considerations is respectfully requested.

Claims 38 to 39 relate to a database, said database containing information on interacting sequence pairs generated by applying the process or method according to the invention. The database represents a similar result as that which is achieved when performing the in-vitro "wet" y-2-h-analysis, namely a compilation of the probabilities of the *in-vivo* interaction of biomolecules.

These "probabilities" may be depicted in one or the other way. The exact wording of the claim reads on a database containing information on interacting sequence pairs generated by applying the process or method according to the invention (representing the result of the method as claimed).

The "database" is claimed by way of a "product-by-process-claim" as the application does not enable the adequate description of the database other than by way of its creation.

The applicant submits, that the examining authority is incorrect when considering the subject matter of claims 38-39 to be covered by Rule 67.1 (v). The applicant requests a reconsideration of this ruling and consequently the formulation of an opinion based on novelty, industrial applicability and inventive step considerations.

The fact that subject-matters excluded from patentability may be involved, or implied, in claim does not render the entirety of the claimed subject matter an activity or subject-matter excluded from patentability.
This means, that although the content of the "data structure" or the "database" needs to be depicted, *e.g.* by means of a screen (= presentation of information) this can not mean that the claimed subject matter is excluded from patentability.

Finally, applicant respectfully submits, by quoting from T 077/92, that "according to the case law established by the Boards of Appeal (before the EPO), an invention is not excluded from patentability under Article 52(2) and (3) EPC if the subject-matter claimed, which has to be assessed as a whole, is technical in character or provides a technical contribution to the prior art, *i.e.* a contribution in a field not excluded from patentability. In decision T 833/91, point 3.1 (not published in the Official Journal EPO) it is stated that "the technical contribution to the art rendering a claimed invention an invention in the sense of Article 52(1) EPC and thus patentable, may lie either in the problem underlying, and solved by, the claimed invention, or in the means constituting the solution of the underlying problem, or in the effects achieved in the solution of the underlying problem."

Taken together, applicant respectfully submits that the above objections are incorrect and kindly requests reconsideration.

B. Re Item IV (Lack of Unity):

The EPO guidelines C III 7.3 state that: "the link between the inventions required must be a technical relationship which finds expression in the claims in terms of the same or corresponding special technical features. The expression "special technical features" means, in any one claim, the particular technical feature or features that define a contribution that the claimed invention, considered as a whole, makes over the prior art. Once the special technical features of each invention have been identified, one must determine whether or not there is a technical relationship between the inventions and, furthermore, whether or not this relationship involves these special technical features."

In view of the above argumentation, applicant respectfully submits that the claims 34, 36-37 and 40-41 are based on a single inventive concept based on the common technical features of inventive methods as claimed in claims 1-31 and the respective claims 34, 36-37 and 40-41, in which the latter represent various forms of products derived from the use of the inventive methods.

Thus, these claims represent a plurality of independent claims in different categories which constitute a group of inventions so linked as to form a single general inventive concept.

C. Re Item V (Rule 66.2 (a) (ii) Reasoned statement with regard to novelty, inventive step or industrial applicability:

Applicant is of the opinion that claims 36-37 and 40-41 are directed to patentable matter, as they are distinguishable over the prior art by means of their technical features as explained in the following.

Claim 36 relates to a computer readable medium for embodying or storing therein data readable by a computer, said medium comprising one or more of the following: a data structure generated by executing a process or a method according to any of claims 1 to 31; computer program code means

which is adapted to cause a computer to execute a process or method according to any one of claims 1 to 31.

Therefore, the technical features are, e.g. computer program code that is characterized by the features which enable the determination of interacting biomolecules, wherein a first group is provided comprising sequences representing homologous biomolecules, at least one second group is provided comprising sequences representing homologous biomolecules, group correlation values between the sequences of the first group and the sequences of at least one second group are determined, and the probability of the interaction of the sequence represented biomolecules is determined on the basis of the group correlation values.

Applicant submits that the same line of argumentation applies likewise to the "computer system" as claimed in claims 40-41.

As outlined in the international application the present invention makes it possible to determine whether or not it is likely that two proteins interact *in-vivo*, analogous to the *in-vitro* yeast-two-hybrid ("y-2-h analysis") analysis, also cited in the international application.

The objective of the "wet" y-2-h-analysis method is "the provision of a method for identifying interacting proteins". The same objective can be formulated for the claimed invention. The efficacy of the claimed invention can be seen for example in the graph in fig. 4 of the present application. This efficiency of the in vitro method of the present invention is surprisingly different from the methods as present in the state of the art and therefore believed to involve an inventive step.

D. Re Item VIII

It is respectfully requested to withdraw the objection with respect to the lack of clarity of the term "biomolecule" as used in claim 34. From the specification, in particular the examples, and fur-

thermore from claim 25, it is clear that the biomolecules of the present invention are DNA sequences, RNA sequences and/or amino acid sequences. Further, the method of the present invention can be readily performed using the information as given on e. g. pages 16 and 21-24 of the specification.

In view of the above argumentation it is respectfully requested to continue with the International Preliminary Examination based on all pending claims 1-41 of the present application and to further acknowledge the patentability of the set of claims as currently on file.

BOEHMERT & BOEHMERT

Christian W. Appelt

Christian W. Appelt